

UNIVERSITY OF HAWAII
LIBRARY

The

Aug 17 '53

THE PHILOSOPHICAL MAGAZINE

FIRST PUBLISHED IN 1798

L. 44 SEVENTH SERIES No. 354

July 1953

A Journal of Theoretical Experimental and Applied Physics

EDITOR

PROFESSOR N. F. MOTT, M.A., D.Sc., F.R.S.

EDITORIAL BOARD

SIR LAWRENCE BRAGG, O.B.E., M.C., M.A., D.Sc., F.R.S.

SIR GEORGE THOMSON, M.A., D.Sc., F.R.S.

PROFESSOR A. M. TYNDALL, C.B.E., D.Sc., F.R.S.

PRICE 15s. 0d.

Annual Subscription £8 0s. 0d. payable in advance

Early Scientific Publications

DIARY OF ROBERT HOOKE, M.A., M.D., F.R.S.

1672-1680



Edited by H. W. ROBINSON and W. ADAMS
Recommended for publication by the Royal Society,
London

25/-

net

"This vivid record of the scientific, artistic and social activities of a remarkable man during remarkable years has too long remained in obscurity."—Extract from foreword by Sir Frederick Gowland Hopkins, O.M., President of the Royal Society.

MATHEMATICAL WORK OF JOHN WALLIS, D.D., F.R.S.

By J. F. SCOTT, Ph.D., B.A.

12/6

net

"His work will be indispensable to those interested in the early history of The Royal Society. I commend to all students of the Seventeenth Century, whether scientific or humane, this learned and lucid book."—Extract from foreword by Prof. E. N. da C. Andrade, D.Sc., Ph.D., F.R.S.
Recommended for publication by University of London

CORRESPONDENCE AND PAPERS OF EDMOND HALLEY

21/-

net

Arranged and Edited by EUGENE FAIRFIELD MACPIKE

First published on behalf of The History of Science Society by Oxford University Press. Now re-issued by Taylor & Francis, Ltd.

MEMOIRS OF SIR ISAAC NEWTON'S LIFE

5/-

net

By WILLIAM STUKELEY, M.D., F.R.S., 1752

From an Original Manuscript
Now in the possession of the Royal Society, London

HEVELIUS, FLAMSTEED AND HALLEY

12/6

net

Three Contemporary Astronomers and their Mutual Relations

By EUGENE FAIRFIELD MACPIKE

Published by arrangement with The History of Science Society

Established
over 150 years

TAYLOR & FRANCIS, LTD.
RED LION COURT, FLEET ST., LONDON E.C.
PRINTERS & PUBLISHERS OF SCIENTIFIC BOOKS

LXXII. *Some Thermal Properties of Point-Contact Germanium Diodes*

By J. R. TILLMAN and J. C. HENDERSON
 Post Office Engineering Research Station, London*

[Received March 25, 1953]

SUMMARY

The isothermal inverse voltage/current relationships of several types of point-contact germanium diode have been measured. They extend well beyond the turnover voltage, which is noted when only slowly increasing currents are applied, but are far from linear. With the assumption, whose validity is discussed, that they can be used to measure the temperature of the barrier layer, other thermal behaviour of the diodes can be investigated. Thus the temperature rise of the barrier layer has been found to be approximately proportional to the wattage dissipated over an important range of wattage; on cessation of a constant dissipation the excess of the temperature of the layer, over the ambient, falls rapidly, e.g. to 40% in about $5\ \mu\text{S}$. The derived data have been used to test a model of the diode, because both the proportionality and the rate of cooling can be correlated with the radius of the hemispherical shell assumed to constitute the barrier layer, the necessary physical constants of germanium being known. Although qualitative agreement is found, there are quantitative differences which are difficult to account for. The data are also used to predict the shape of the steady state relationship up to and beyond the turnover voltage. Here good agreement is seen, which is taken as confirming the idea that turnover is largely the result of self-heating.

§ 1. INTRODUCTION

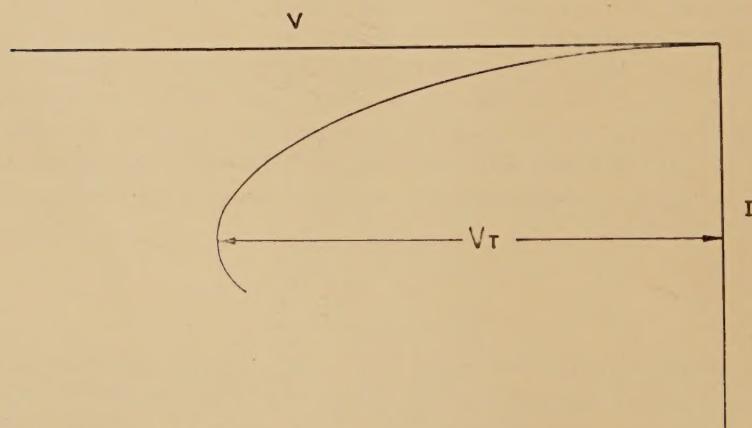
THE relationship between current, I , and voltage, V , of commercial† point-contact germanium diodes can conveniently be divided into three parts. The first part occurs in the voltage range $+0.1$ upwards (whisker

* Communicated by the Authors.

† The electrical properties of the contact between a metal point and a block of germanium are often both improved and made stable with time by forming—a process which involves, momentarily, a high rate of dissipation of heat at the contact in the presence of an electric field, usually intense. There is no unique forming process. Scientifically, forming is not well understood. It probably increases the contact area, but little is known quantitatively of the modifications it causes to the germanium in the immediate vicinity of the contact. However, it does not follow that our ignorance of the mechanism of rectification is necessarily increased by forming for there are many gaps in our knowledge of the state of the surface of the germanium before forming takes place, which may be less important after forming. 'Commercial' implies not only that the units are offered for sale, but that they have most probably been formed.

positive with respect to the n-type germanium), the second in the range $+0.1$ to -0.1 v and the third for inverse voltages greater than 0.1 . The first and third parts merge smoothly with the second. The relationship in the first range of voltage is important in practice for several reasons ; it largely determines the rectification efficiency in many applications and the scale shape of meters using the diode. It has proved more complex than any theoretical relationship based on spreading resistance, which is a function only of the size and shape of the contact and of the bulk resistivity of the germanium. The relationship in the second range is of importance in the detection of low-level signals and enables some tests of theory to be made. Its temperature dependence should be related to the height of the potential barrier at the contact (Billig and Ridout 1951). The relationship in the third range, to which this paper is confined, is both of much practical and theoretical importance. It may determine the maximum signal level that can be handled, and the efficiency of units when blocking a.c. signals or direct current. It imposes severe tests on theories of the mechanism of rectification and the location and size of the rectifying layer.

Fig. 1

Typical inverse V/I relationship measured with direct current.

Among the more important features of the inverse relationship is that of the maximum (turnover) voltage V_T which occurs as the inverse current is increased from zero (see fig. 1). V_T is not only a function of the make up of the diode (see e.g. Douglas and James 1951) but is also a function of ambient temperature, and, as will be shown, of the rate at which the current is increased, if fast. Benzer (1949) made an extensive study of V_T and of the power W_T dissipated at the maximum reverse voltage sustainable. He showed that W_T fell linearly with increase of base temperature θ_b and suggested that the contact temperature at the peak voltage was approximately independent of θ_b .

As a result of Benzer's work the turnover of the V/I relationship began to be regarded primarily as a thermal effect, but much remained to put this view on a sounder footing. Hunter (1951) took an important step when he suggested that this part of the relationship can be "explained entirely on the basis of self-heating if one assumes: (1) that the isothermal characteristics are straight lines [passing through the origin, and measurable with pulses of so short duration and low repetition frequency that self-heating is negligible]; (2) . . . that the difference between the temperature immediately beneath the point and the ambient temperature is directly proportional to the power dissipated; (3) that the conductivity of the isothermal characteristic shows a thermal activation energy, $e\phi$, of the order of one electron volt [$e\phi=0.76$ ev was assumed later]". He went on to show that the V/I relationship measured with d.c. could then be predicted, qualitatively, out to and beyond turnover, although no mention or justification was made of the very important constant of proportionality adopted between increments of power dissipated and increments of temperature resulting.

Bennett and Hunter (1951) simultaneously reported measurements of isothermal characteristics and compared them with the V/I relationship measured with d.c. They found that (a) the duration of the pulses needed ought not to exceed $0.5\ \mu\text{s}$, (b) that the isothermal characteristics were in fact nearly straight lines out of the origin to voltages of more than $2V_T$, (c) that these lines were separated for increasing values of temperature, T , by much less than Hunter's third assumption [$I/V \propto \exp(-e\phi/kT)$ with $e\phi=0.76$ ev] implied and (d) the currents passed at voltages from, say, $0.1V_T$ to, perhaps, V_T were, for values of θ_b about 20°C , sometimes less when measured with d.c. than with narrow pulses. The disproving of Hunter's third assumption was not important qualitatively, but finding (d) did upset Hunter's suggestion severely. In order to account for the ratio V/I obtained from d.c. measurements being greater than that from isothermal measurements, Bennett and Hunter turned to Aigrain's theory (Aigrain 1950) in which the rectifying barrier is assumed to be underneath a thin surface layer whose resistivity, though relatively low, can have a positive temperature coefficient in the region of room temperature. Although a qualitative explanation of their results was then possible, it could not easily be subjected to quantitative tests.

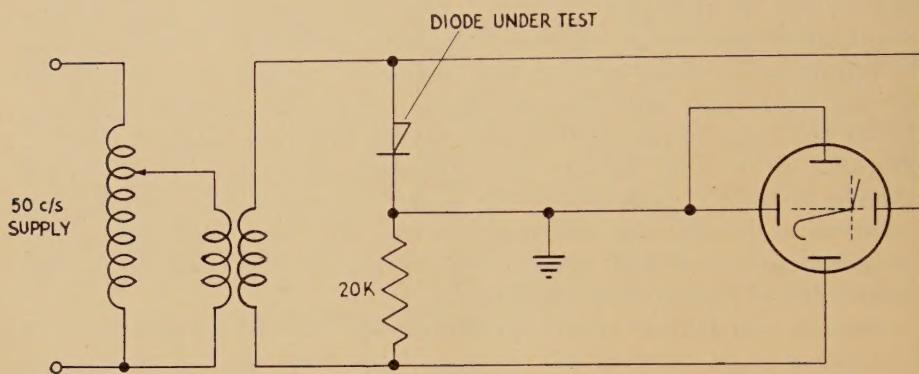
§ 2. EXPERIMENTAL METHODS AND RESULTS

2.1. Preliminary

The measurements now to be described arose out of some made two years ago on the dependence of V_T on the conditions of measurement. The use of alternating voltages (say at 50 c/s) had long been known to make measurements easily presentable on an oscilloscope (see fig. 2) and to give results only very little different from those obtained using d.c., but some brief studies have been made using higher frequencies. Thus Benzer found that the narrow loop which is so often observed at 50 c/s

(see the full curve of fig. 3) widened as the frequency was increased up to 5 kc/s, the highest he used. Henisch and Granville (1951) while investigating metal-galena contacts found that V_T increased as the frequency was raised to 10 kc/s, their highest. Measurements at frequencies, f , above 10 kc/s present new but not severe difficulties.

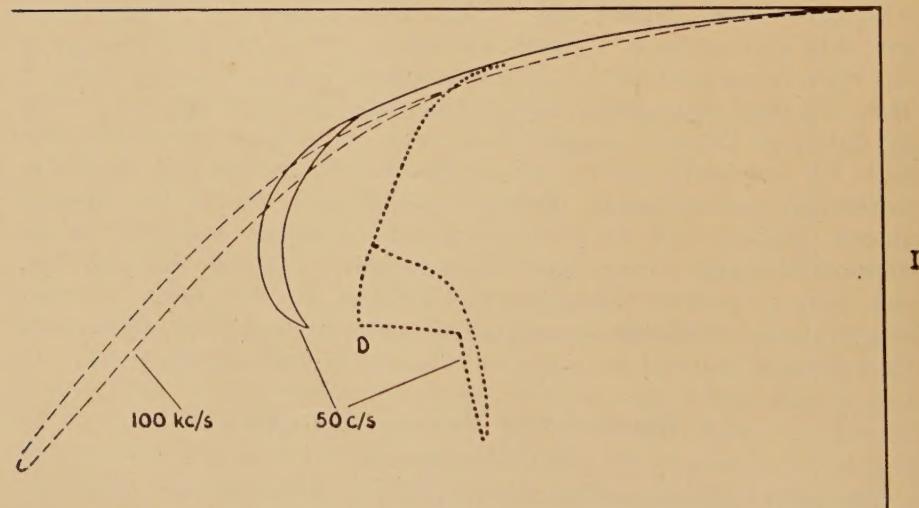
Fig. 2



Apparatus for observing V_T with an alternating supply of low frequency.

Fig. 3

V



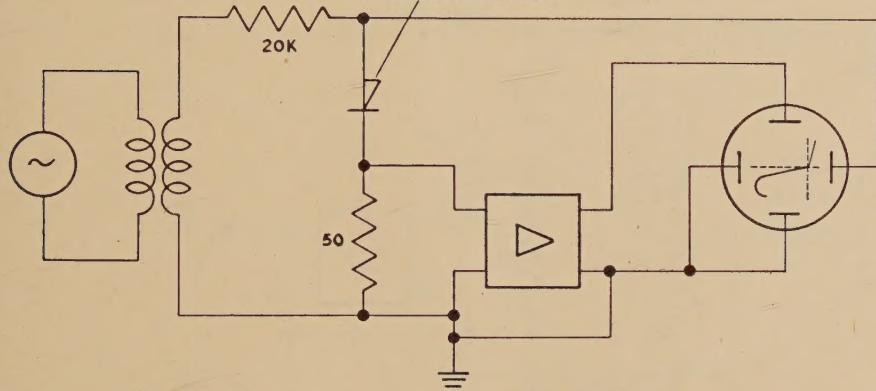
Typical inverse V/I relationships measured with alternating current.

Phase shifts introduced by either the earth capacitances (e.g. of the feeding transformer) in fig. 2 or the amplifier needed in the preferred circuit of fig. 4, cause a loop to appear even when a resistor is substituted for the diode, and must be made negligible at frequencies from f to

several times f , because the waveform of the current passed by the diode contains important components at the lower harmonic frequencies of f . Apparatus suitable for use at 100 kc/s has been made; it shows that V_T then loses all significance with many diodes (see the broken curve of fig. 3 which was observed for the unit giving the full curve at 50 c/s). The result is consistent with the view that V_T arises from self heating, provided the thermal time constant of the rectifying layer is not much less than 1/100 000 sec, a value suggested by Benzer from some elementary experiments. Consideration was therefore given to a study of the V/I relationships for a number of values of f , with a view to obtaining a value of the time constant. After some preliminary measurements on these lines however—during which it was noted that, for some frequencies, the portion of the curve outgoing from the origin and the portion incoming crossed one another for a few units—another method, described in

Fig. 4

DIODE UNDER TEST



Apparatus for observing V_T with an alternating supply of high frequency.

§ 2.4 was found to be preferable. Benzer had also noted shapes of V/I relationships other than that of the full-line curve of fig. 3; one in particular, that shown by the dotted curve of the same figure was noted frequently in measurements at 50 c/s with the types of diode now tested. When it applied it caused complications in some later measurements and deductions.

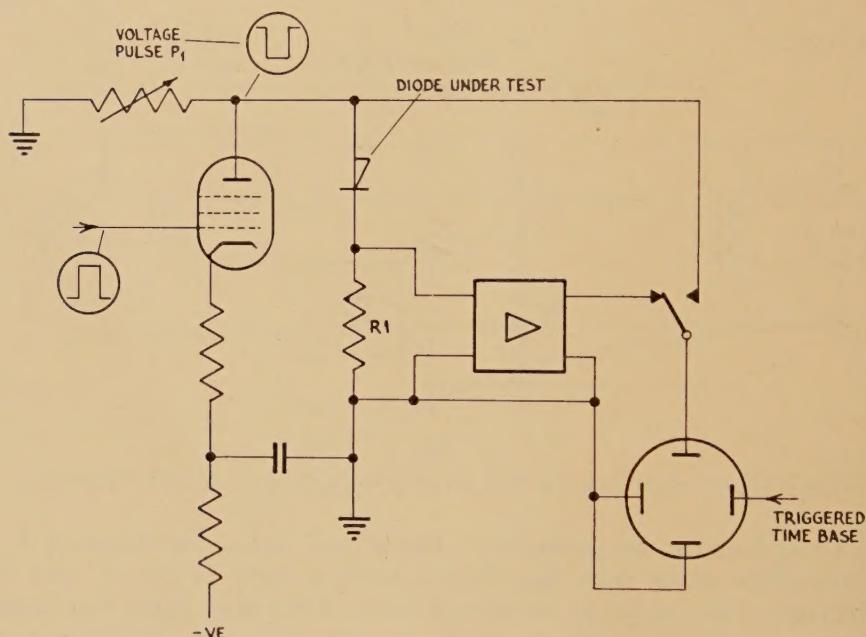
Measurements similar to those described by Hunter were also made, followed by a test of the validity of assuming a constant of proportionality C , between temperature rise, θ , of the barrier layer and the steady dissipation of electrical power $W (= VI)$, thus $C = \theta/VI$. A knowledge of C and the isothermal characteristics can then be used to predict the d.c. characteristic. Moreover, because C must be a function of at least one dimension of the barrier layer and of the thermal properties of germanium, it enables an estimate of the dimension to be made. A further experiment has led to the deduction of the relationship between

temperature of the barrier and time after the cessation of a known power dissipation and hence to a second estimate of the same dimension. In the interpretation of the measurements which follow, the commonly-adopted approximation is made that the whole of the inverse voltage applied to, or sustained by, the diode falls across the barrier layer, although in fact a small proportion, often no more than 1%, falls elsewhere.

2.2. The Isothermal Characteristics

The measurement of the isothermal relationship between V and I presented no difficulties with the apparatus shown schematically in fig. 5. The amplifier necessarily employed responded rapidly (within $0.1 \mu\text{s}$) to changes of input voltage and provided an adequate current sensitivity,

Fig. 5

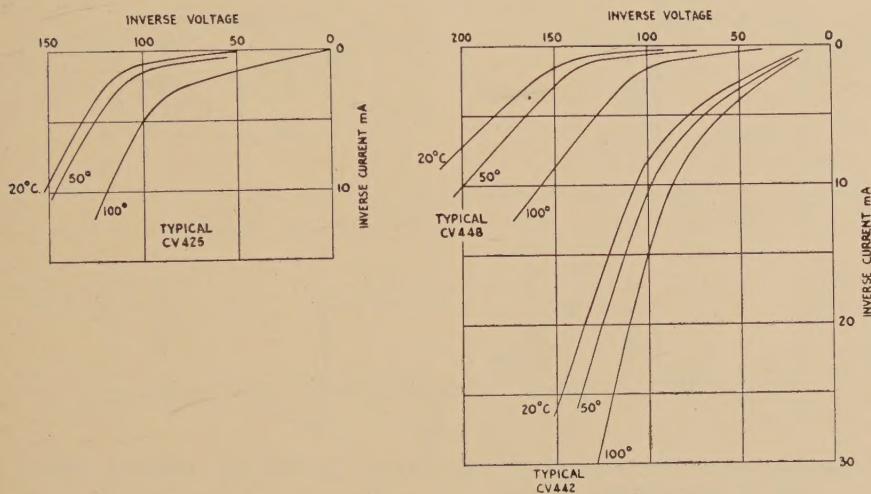


Apparatus for measuring isothermal, V/I , relationships.

e.g. 1 mA through R_1 ($= 50\Omega$) deflected the spot 1 cm over the face of the cathode ray tube. The test pulses P_1 had steep sides and, for most tests, a repetition frequency of about 1 kc/s and a duration of about $1 \mu\text{s}$ —no necessity was noted for pulses shorter than $0.5 \mu\text{s}$ for the large majority of diodes tested. The diode was contained in a small oven heated by a stream of hot air; the temperature-sensitive element of the oven's thermostat was the exposed filament of a small lamp, whose small thermal capacity and time constant reduced the time required to stabilize the temperature of the oven.

Only diodes classed as free from the effect noted by Meacham and Michaels (1950) (the passage of a large transient current in the reverse direction on the application of a reverse voltage immediately following the passage of a forward current) were tested. Typical results obtained from diodes meeting the specifications for CV 425, CV 442 and CV 448 are shown in fig. 6. When those applying at temperatures up to at least 50°C are compared, diode by diode, with the results obtained with the apparatus of fig. 2 and the same ambient temperature, the currents passed, at voltages from $V_T/3$ to $2V_T/3$ at least, were slightly greater under the isothermal conditions for many of the units—the unexpected behaviour noted by Bennett and Hunter. The sensitivities of the two measurements were inadequate for firm comparisons to be made at voltages less than about $V_T/3$, for which the currents passed were often <0.1 mA.

Fig. 6

Typical isothermal V/I relationships.

For a few units the isothermal characteristics, in the range of temperature 20–50°C, changed in the opposite sense to that shown in fig. 6, i.e. less current was passed for a given amplitude of the pulse voltage as the temperature was raised, making the unexpected behaviour noted above capable of qualitative explanation.

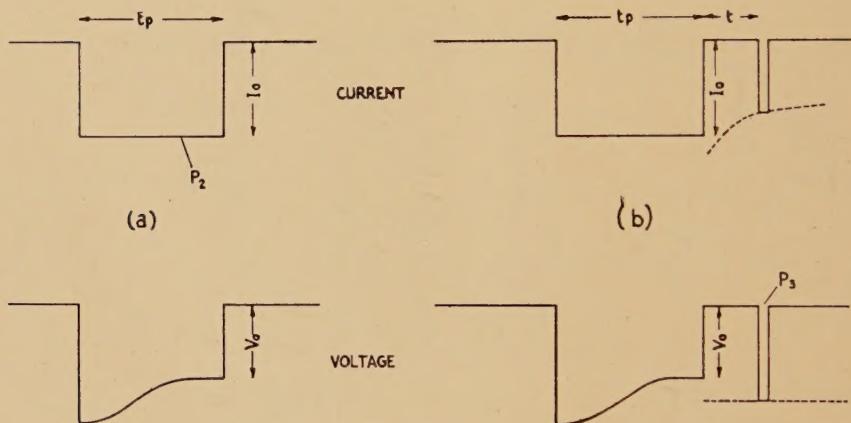
The curves of fig. 6 depart markedly from straight lines well before a voltage of $2V_T$ is reached, in contrast to those of Bennett and Hunter.

2.3. The Proportionality between Power Dissipated and Temperature Rise

The basis of the test for, and the measurement of, the proportionality between the power dissipated, $W = VI$, and temperature rise, θ , is simple if the isothermal characteristics can be used as a measure of temperature.

Suppose the rectifying layer is heated by the passage of a constant current I_0 in the reverse direction for a time t_p sufficient to reach thermal stability (see fig. 7 (a)). Then if the ultimate voltage V_0 set up across the barrier by I_0 is measured, reference to the isothermal characteristics will give the temperature θ_h of the layer, proper to the dissipation $V_0 I_0$. For ease of measurement the test apparatus (shown schematically by the circuit of V1 of fig. 8) uses a train of current pulses (P_2); the train gives rise to some bulk heating of the diode, and causes the temperature, θ_c , of the barrier immediately prior to each pulse of current to be slightly above the ambient. However, because $200 \mu\text{s}$ always proved sufficient as t_p and a low frequency (200 c/s) of pulse repetition could be used without difficulty, it was confidently estimated that θ_c never exceeded the ambient temperature by more than 2°C —often by much less. The proportionality sought is $C = (\theta_h - \theta_c) / V_0 I_0$.

Fig. 7



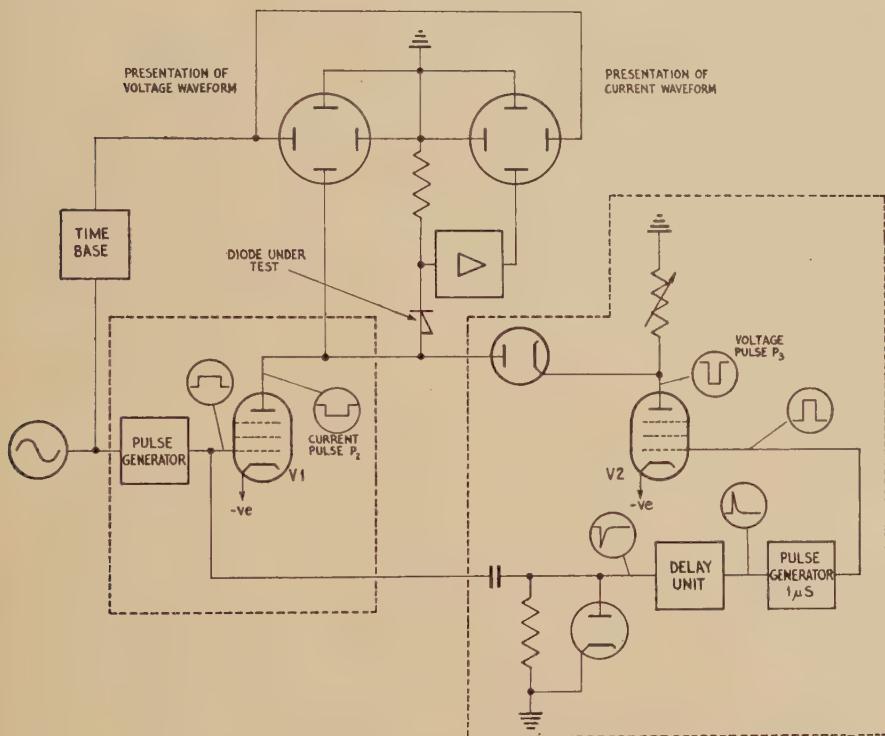
(a) Waveforms of the pulse of reverse current P_2 and the resulting voltage; used in determining C .
 (b) Waveforms when P_3 is also in use.

Whereas the measurements of the isothermal characteristics could be made repeatedly without either the results, or any other property of the diode, changing, the measurements using the train of heating pulses, of current about 10 mA, caused the isothermal characteristics to change considerably for some types of diode. The use of less energetic heating pulses was tried, but because, in general, these types of diode did not show widely spaced isothermal characteristics, the sensitivity of the experiments was poor. Only types able to withstand the higher currents during heating were investigated in detail.

The dependence of C on the product $V_0 I_0$ is shown in fig. 9 for four units. The discontinuity in the curve for unit (b) was typical of units showing V/I relationships like that of the dotted curve of fig. 3. The other curves of C were obtained for units showing V/I relationships as

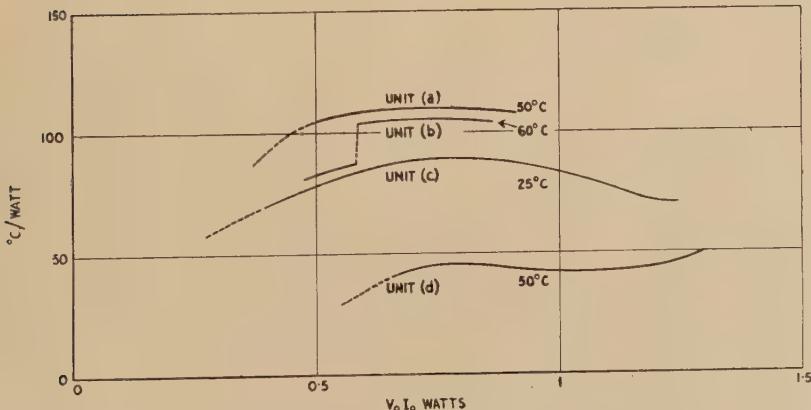
the full curve of fig. 3 ; the isothermal relationships of the unit (b) were not qualitatively distinguishable from those of units (a), (c) and (d) however.

Fig. 8



Apparatus for the measurement of C and of cooling curves.

Fig. 9



The dependence of C on $V_0 I_0$.

The temperature marked against each curve is the ambient used during its measurement.

The sensitivity of the experiments fell as $V_0 I_0$ was reduced and values of C for $V_0 I_0 < 0.3$ w are much less accurate than those for higher wattages; there were strong indications however that C fell sharply as $V_0 I_0$ was reduced below 0.25 w.

2.4. The Relationship between the Temperature of the Barrier Layer and Time, on the Incidence and Cessation of the Dissipation of Power in the Layer

The isothermal characteristics can also be used to convert the decay of the voltage V set up across the diode, in the early stages of the incidence of the constant reverse current I_0 , to a relationship between temperature (θ) and time (t). The results obtained are measures of thermal behaviour and would seem useful in comparing models of diodes, the necessary physical constants of germanium now being known. However, although the current passed is constant with time, the power dissipated ($W = VI_0$) is neither constant nor a simple function of time, making analytical predictions of the θ/t relationship very difficult, even for the simplest models. As is common in the prediction of transient behaviour, a step function, i.e. no dissipation for $t < 0$ and a constant dissipation for $t > 0$ or vice versa, proves most amenable to analysis. The control of the relationship between I and t to give a constant value of VI would be very difficult in practice however and was not attempted because the equivalent data should be obtained more simply by the addition of the apparatus shown schematically by the circuit of V2 in fig. 8. The cooling of the unit, on the cessation of the constant dissipation $I_0 V_0$ (I_0 having been applied for long enough to stabilize V_0), is then investigated with a pulse of constant voltage, P_3 . The pulse, in practice one of a train of pulses, is of short duration (e.g. $1 \mu\text{S}$, as for P_1) so as not to disturb the cooling and is delayed by a time t controllable with respect to the cessation of I_0 (see fig. 7 (b)). The temperature θ at t is deduced from the amplitude of the pulses of current passed on the application of P_3 , with the aid of the isothermal characteristics. θ_c was derived indirectly in this way. An ambient (oven) temperature of $\sim 50^\circ\text{C}$ was commonly used in these experiments because the isothermal characteristics at lower temperatures are not always sufficiently spaced to give the experiments the sensitivity required.

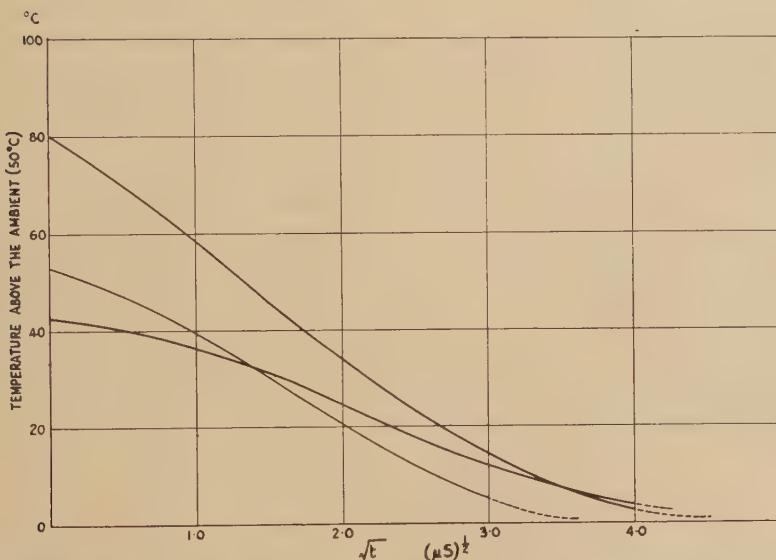
Typical relationships between θ and t are shown in fig. 10 for three units. The ordinates at time $t=0$ differ because the powers dissipated differed and the units had different values of C . The reason for using \sqrt{t} as the abscissa will be seen later. The rapid fall of θ with t during the first μS or so would, on its own, suggest that the pulse duration required for accurate measurement of the isothermal characteristics is $< 0.5 \mu\text{S}$. On the other hand, it was noted that the fall of voltage (and hence the rise of temperature) occurring initially during the application of the pulse of current I_0 was much slower than the fall of current (and hence the fall of temperature) observed with the pulses of voltage P_3 . This finding, suggesting that heating and cooling are not so exactly complementary, must have a bearing on the deductions drawn from the measurements.

§ 3. DISCUSSION

3.1. Assumptions Involved in the Use made of the Isothermal Characteristics

Before attempting to make deductions from the quantity C (and from the relationships between temperature and time shown in fig. 10), the assumption—that the isothermal characteristics can be used to measure temperature in the experiments described in §§ 2.3 and 2.4—must be considered. It implies that the pulse of current I_0 raises the temperature of the barrier layer uniformly throughout, as does heating in an oven. There are several reasons however for believing that uniformity may not necessarily be achieved. In particular, because the layer may differ in conductivity over its area, patches of high conductivity will carry more than their share of the total current, will dissipate

Fig. 10



The decay of barrier temperature with time, t , for three units following a temperature rise produced by the dissipation of power at the barrier.

more than their share of the total power and will therefore rise in temperature more than the average. The extra rise in temperature tends to accentuate their higher conductivity and hence the patchiness. In addition, despite what is assumed at first in the next section, the thermal environment of the barrier is not uniform.

Furthermore the assumption is made that the mechanism of conduction applying initially (e.g. in the first microsecond) on the application of a pulse, such as that of I_0 , is that applying at any later time. If, for instance, minority carriers or a trapping process plays a part in the mechanism, time-dependent conduction may result from other than temperature changes.

3.2. Deductions to be Drawn from Thermal Considerations Only

Before dealing with what was to have been the main purpose of the investigation—the prediction of the d.c. characteristics from more fundamental data—the thermal behaviour of a simple model of the diode will be considered in some detail and compared with the data for C and with fig. 10.

A point source of heat, generating H cal/sec, in an infinite homogeneous solid of thermal conductivity k , specific heat σ and density ρ ultimately sets up a temperature/space relationship $\theta = H/4\pi kr$, where r is the distance from the point, and θ is the rise in temperature above the ambient (i.e. above the temperature at infinity). If H ceases at time $t=0$, the relationship between temperature, θ_1 , and radius, r , at any subsequent time, t , is $\theta_1(r, t) = H/4\pi kr \operatorname{erf} r/\sqrt{4Kt}$, where $K = k/\rho\sigma$; the ratio $\theta_1(r, t)/\theta_1(r, 0)$, where $\theta_1(r, 0) = H/4\pi kr$ applies at $t=0$, is shown by the full line of fig. 11 in terms of the variable \sqrt{Kt}/r . The values of k , ρ and σ for germanium are $0.14 \text{ cal } ^\circ\text{C}^{-1} \text{ cm}^{-1} \text{ sec}^{-1}$ at 25°C falling to ~ 0.11 at 100°C ,* 5.5 g cm^{-3} and 0.074 respectively. A hemispherical model with no radiation losses can be treated similarly and the thermal behaviour of an idealized barrier layer—a shell of radius r_0 (and thickness $\ll r_0$) predicted.

Several objections can be raised to a hemisphere, with a point source of heat at its centre, being taken as a model. Some attempt must be made to assess the significance of the more important objections. Thus consider, as a better approximation to practice, that the heat is generated in the shell of radius r_0 and thickness small compared with r_0 . For the infinite solid the steady state relationship between θ and r remains unchanged for $r \gg r_0$, but θ becomes independent of r when $r < r_0$, having the constant value $H/4\pi kr_0$. The heat contained inside the shell at $t=0$ is therefore $Hr_0^2\rho\sigma/3k$ instead of

$$\int_0^{r_0} \frac{H\rho\sigma}{4\pi kr} \cdot 4\pi r^2 dr = Hr_0^2\rho\sigma/2k.$$

Of more importance perhaps is the changed relationship between the temperature θ_2 of the shell, after the cessation of H , and time; it is

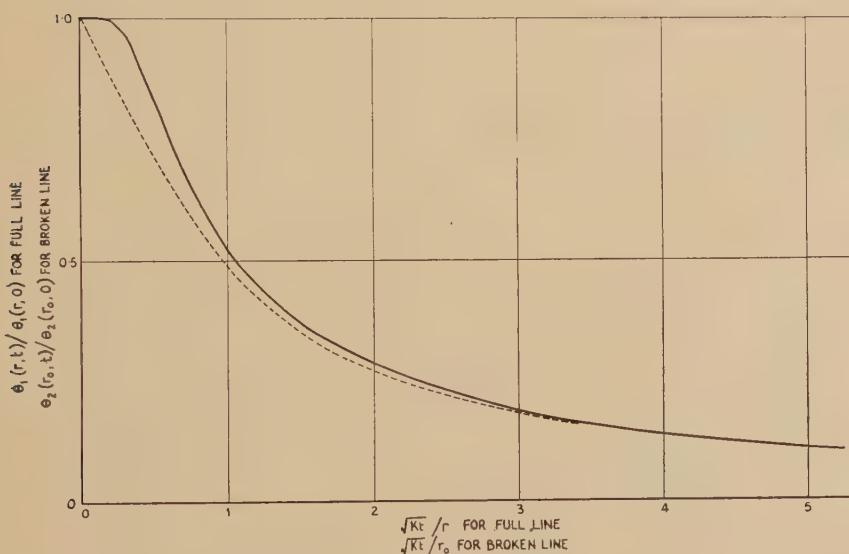
$$\theta_2(r_0, t) = \frac{H}{4\pi kr_0} \left[\operatorname{erf} \frac{r_0}{\sqrt{Kt}} - \frac{1}{r_0} \sqrt{\left(\frac{Kt}{\pi}\right)} \{1 - \exp(-r_0^2/Kt)\} \right].$$

$\theta_2(r_0, t)/\theta_2(r_0, 0)$ is shown by the broken curve of fig. 11 as a function of \sqrt{Kt}/r_0 . The two curves differ for small values of \sqrt{Kt}/r_0 but come together for large values (r_0 is the only value of r for which the two curves can be compared). The differences resulting from the change of model are not greater than the experimental error of the results shown in fig. 10.

* In subsequent calculations a value of 0.12 has been taken as perhaps the best mean for temperatures 50° – 100°C .

The new model, suitably modified to be hemispherical, must in addition take the metal whisker into account. The thermal conductivity, k , and the product, $\rho\sigma$, of the metal of the whisker, tungsten, are respectively 3.3 times and 1.6 times those of germanium, making K 2.0 times. But the solid angle which the whisker subtends is not easily assessed. The taper of the whisker suggests an angle no greater than $\pi/10$, in comparison with 2π subtended by the germanium. But considerations of the immediate surroundings of the barrier layer suggest that the angle effective for the purpose of assessing cooling may be greater.* If, for the present, the effective angle is given the value $x2\pi$, the expression for $\theta_2(r_0, t)$ is perhaps sufficiently modified by multiplying k by $(1+3.3x)$ and K by $(1+2x)/(1+x)$, k and K retaining the values appropriate to germanium.

Fig. 11



Theoretical cooling curves for two models.

The first factor takes into account the higher thermal conductivity of tungsten in assessing the total solid angle expressed as if it were entirely

* An approximate numerical analysis (using relaxation methods) has been made by W. E. Thomson of the flow of heat along a whisker. The heat is supplied over the surface of a hemisphere at constant temperature, corresponding to the portion of the whisker embedded in the germanium; the rest of the whisker is represented by an infinite cone of solid angle $\pi/10$. The results show that with the whisker and the germanium having the same thermal conductivity and constant temperature on the hemispherical boundary, the ratio of heat flows is three times the ratio of solid angles; i.e. the effective solid angle of the whisker is $3\pi/10 = 0.15 \times 2\pi$.

of germanium; the second factor takes account of the relative importance of the germanium and of the whisker in deriving a weighted value of K .

More complex models, taking into account finer points of the make-up of the practical diodes, e.g. the finite size of the block of germanium, the presence of the base support and the atmosphere in contact with the whisker and block, do not lend themselves to analytical solutions. They are not pursued because they do not seem likely to result in large quantitative changes to results based on expressions already given.

Deductions will therefore be drawn from the experimental data only in terms of the model developed so far.

A comparison of the results given in §§ 2.3 and 2.4 with the expression deduced for $\theta_2(r_0, t)$ enables two values of r_0 , designated r_{01} and r_{02} , to be obtained in terms of x . Firstly, because $\theta_h - \theta_c$ can be identified with $\theta_2(r_0, 0)$ and H with $V_0 I_0 / J$, where $J = 4.18 \text{ cal watt}^{-1} \text{ sec}^{-1}$,

$$r_{01} = \frac{1}{2\pi J(1+3.3x)kC},$$

where $C = (\theta_h - \theta_c) / V_0 I_0$. Secondly if a curve of fig. 10 can be made approximately to coincide with the broken curve of fig. 11 by multiplying the \sqrt{t} axis of the former by a , r_{02} is given by

$$\frac{1}{a} \sqrt{\left[\frac{(1+2x)}{(1+x)} K \right]}.$$

In practice a fair fit can be made from $\theta_2(r_0, t) / \theta_2(r_0, 0) = 1$ to about 0.4 and accordingly only those parts of the curves of fig. 10 have been used in deducing r_{02} .

The table shows results obtained for r_0 for three diodes, both in terms of x and with x given two specific values, 0.1 (which might well be less than the true value) and 0.5 (which may well be several times the true value). Some knowledge of r_0 can also be obtained by visual inspection of units of the type tested, because r_0 must be greater than the radius of the approximately hemispherical end of the whisker. The lower limit seems, from an inspection, to be perhaps as small as 0.0007 cm, with which no value of r_{01} or r_{02} conflicts. The general agreement between the values of r_{01} and r_{02} deduced, from C and the cooling curves respectively, is so poor however, as to cast doubt on the model.

Patchiness of the barrier layer can not be invoked to explain the discrepancy between r_{01} and r_{02} . It should lead to high values of C and hence to low values of r_{01} , and although it might also lead to a more rapid fall of θ with \sqrt{t} in fig. 10 than the model suggests and hence to low values of r_{02} also, it is unlikely to do so sufficiently to effect agreement between r_{01} and r_{02} even if the true value of r_0 is several times any value quoted so far and the patchiness is very marked. On the other hand, an estimation of the θ /time relationship during the pulses of I_0 (see fig. 7) suggested a time scale greater than that shown in fig. 10 and hence larger values of r_{02} .

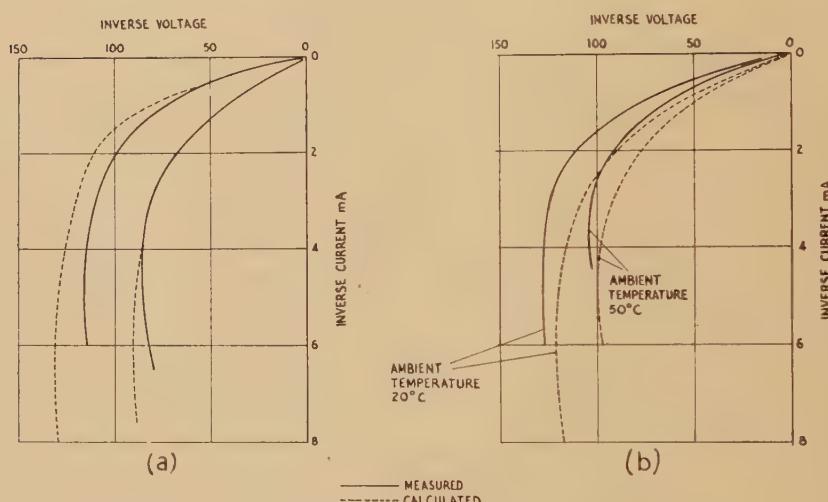
Values of r_0 deduced from C and a

Diode No.	$C = \frac{(\theta_h - \theta_c)}{V_0 I_0}$ °C/watt	$r_{01} (1 + 3x)$ $= \frac{1}{2\pi k C J}$ cm	$r_{02} \sqrt{\frac{1+x}{1+2x}}$ $= \frac{\sqrt{K}}{a}$ cm	$x = 0.1$		$x = 0.5$	
				r_{01} cm	r_{02} cm	r_{01} cm	r_{02} cm
1	108	2.95×10^{-3}	630	0.87×10^{-3}	2.2×10^{-3}	0.90×10^{-3}	1.1×10^{-3}
2	78	4.1×10^{-3}	670	0.82×10^{-3}	3.1×10^{-3}	0.85×10^{-3}	1.55×10^{-3}
3	44	7.2×10^{-3}	500	1.1×10^{-3}	5.4×10^{-3}	1.15×10^{-3}	2.7×10^{-3}

3.3. Deduction of the Steady State Relationship between Voltage and Current and hence of the Turnover Voltage

A knowledge of the isothermal characteristics of a unit and of its value of C (if constant over a sufficient range of I_0) enables prediction to be made of the steady state relationship between voltage and current, every point (V, I) of which must lie on that isothermal characteristic whose temperature (θ above the ambient) also satisfies the equation $\theta = CVI$. Figures 12 (a) and (b) show curves constructed from series of points obtained in this way, compared with those measured directly, for units having characteristics similar to that of the full curve of fig. 3. The value of C used for each unit was a mean taken from a curve similar to those shown in fig. 9. An ambient temperature of 50°C was generally used, because the opening-out of the isothermal characteristics above

Fig. 12



Measured and Calculated steady-state V/I relationships.

- (a) Two typical units. Ambient temperature 50°C .
- (b) A third unit at two ambient temperatures.

that temperature increases the sensitivity. Calculations made for a few units over a range of ambient temperature from about 20°C (one curve is shown in fig. 12 (b)) up to at least 80°C showed that the calculated turnover voltage fell as the ambient temperature rose, agreeing, approximately quantitatively, with practice. There was no evidence to suggest however that the voltage and current applying at turnover lay exactly on one particular isothermal relationship independently of the ambient temperature used.

It is doubtful whether the comparison between each pair of curves (calculated and measured) is worth close examination. Certainly too much should not be read into the good quantitative agreement sometimes

noted because, providing V_0 is the steady state voltage for the current I_0 , the steady state characteristic can be constructed from the measurements of V_0 over a range of values of I_0 without recourse to the deduction, and subsequent use, of C . The agreement has however been achieved by the use of a mean value of C for each unit, although experiment indicates that C falls off for small values of I_0 —a factor which would be expected to influence the V/I relationship up to about the turnover point. Of more importance than good quantitative agreement is the fact that turnover appears in the calculated curves.

The V/I relationships were also calculated for some units whose measured characteristics resembled that dotted in fig. 3. Two values of C were used, the first appropriate to wattages less than that at which the discontinuity in C occurred (see the curve for unit (b) of fig. 9) and the second to greater wattages. The calculated relationships agreed well with the measured, both up to and beyond the point D of fig. 3, but they contained nothing to explain the large loop observed. The loop is characterized by the fact that it changes little as the frequency of the a.c. used to display it is increased from 20 c/s to 1000 c/s. Hence it seems more a phenomenon of electrical hysteresis than one due to a time constant of the diode (thermal or otherwise).

3.4. General

The inverse V/I relationships of point-contact diodes have proved too complex to be quantitatively explained by the simple model adopted. Nonetheless several properties have been more fully reported than hitherto and the deductions drawn from them used to test earlier tentative statements about the relationship. The isothermal characteristics, though extending well beyond the turnover voltages observed for d.c. measurements, are not straight lines and show an activation energy well below that of germanium, 0.72 v. Their departure from straight lines, though not sudden, takes place at a voltage ~ 100 v at room temperatures for many of the units tested.

They can be compared with Gunn's suggested two-part relationship (Gunn 1952). The first part, $I \propto V$, applies when $I < I_1 = 2\pi r_0^2 \sigma_0 E_1$, where $E_1 (= 1.7 \times 10^3 \text{ v cm}^{-1})$ is the field at which the drift-velocity of electrons ceases to increase with increase of field and σ_0 is the electrical conductivity of the germanium used; the second part

$$I = \frac{I_1}{E_2^2 r_0^2} (V + E_2 r_0 - V_0)^2$$

applies for greater currents, where E_2 is the field necessary to produce the Zener effect ($\sim 2 \times 10^5 \text{ v cm}^{-1}$) and V_0 is the drop across the barrier layer, implied by Gunn to be relatively unimportant. Because σ_0 probably lies between $0.1-0.5 \Omega^{-1} \text{ cm}^{-1}$ and $r_0 \sim 0.001 \text{ cm}$, I_1 probably lies between 1 and 5 mA. Without knowledge of V_0 it is not possible to say at what value of V the second part should take over, but because $r_0 E_2 \sim 200$ v, I should not reach, say, $4I_1$ until V exceeds V_0 by at

least 200 v. Quantitatively therefore the curves of fig. 6 show little agreement with Gunn's theory, but in view of his simplifying assumptions and the fact that the diodes now tested have been formed, closer comparison is not justified.

They can also be compared with the expressions developed by Simpson and Armstrong (1953) for the inverse characteristics, who took account of three components (including hole concentration) determining the field at the contact. In the expressions evolved for the current, both $\log I \propto V^{1/2}$ and $\log I \propto V^2$ arise, but neither fit the curves of fig. 6 well. Radii, other than r_0 , are introduced by Gunn, and Simpson and Armstrong. The radii are voltage- and temperature-dependent, though differently so for their two models, and may influence the measurements of §§ 2.2, 2.3 and 2.4 differently. Thus the radius inside which most of the heat is dissipated, the radius from which the carriers are drawn to give the current which is used to measure the temperature, and that of the region whose conductance determines the isothermal characteristics, may differ significantly. The fall of C at low values of $V_0 I_0$ may arise from some of these differences.

The proportionality between temperature rise θ of the barrier layer (across which the majority of any inverse voltage applied has been assumed to fall) and the power $V_0 I_0$ dissipated has been shown to be substantially constant over a range of power for some units, though doubt exists at low powers. The measurement of proportionality has made use of the isothermal characteristics, which is almost certainly not fully justified. Other methods of measurement which have been suggested seem also to depend on assumptions—e.g. that the thermal state of the barrier layer is the same at the point of turnover independently of the ambient temperature—which are equally open to criticism. If, as seems probable, the temperature of the barrier is not, in general, uniform when a steady state is reached, it may prove very difficult to make accurate deductions from any measurements.

The work on the proportionality between θ and $V_0 I_0$ has led to an estimate of the temperature rise likely during normal use. The rise is small (rarely exceeding 70°C even at turnover) and, on its own account, is unlikely to cause any permanent physical change in the diode, due to diffusion of impurities or lattice imperfections, even if maintained for several years. Of more importance is the extent to which it assists the high electric field accompanying it to bring about changes, a subject which, it seems, has yet to be investigated.

The transient thermal response of the diode, deduced once again with the aid of the isothermal characteristics—with whatever inaccuracies that deduction entails—illustrates several properties of the diode. The term 'time constant' has been avoided because it is not strictly applicable; users will however wish to know for what durations a particular voltage can be sustained in the reverse direction without the current increasing by more than a given fraction and possibly a given

current passed without the voltage sustained falling by more than a given fraction. Curves such as those from which fig. 7 was drawn and those of fig. 10 will help them to decide. The time scale observed is generally in fair keeping with the reciprocal of the frequency at which turnover loses its significance when observed with the apparatus of fig. 4.

The deduction of r_0 from both C and the temperature/time relationships is seen, on close scrutiny, to involve several approximations which are difficult to assess. Here particularly it does not seem likely that a fully coherent picture can be obtained without taking into account features of the model which make analysis difficult. Moreover the present experimental accuracy may need to be improved, e.g. in the measurement of C at low powers, and of the θ/t relationship, before all the implications of any model can be sufficiently assessed.

The steady state V/I relationship has been deduced from more fundamental data by assuming that the barrier layer heats up uniformly when dissipating power and that the constant of proportionality C holds over the relevant range of VI involved—two suppositions which are known not to be fully justified. Even so the agreement with the measured relationship is so close as to make it difficult to believe than any mechanism other than self-heating can be that primarily responsible for the turnover of voltage.

§ 4. CONCLUSIONS

Although the investigation has provided new data of the inverse voltage current relationship of germanium diodes, the attempts to fit the data to a model have met with only partial success. While, qualitatively, reasons can be advanced to explain the important discrepancies outstanding, it is not easy to design experiments to support the reasons. The primary deductions, on which most reliance rests, conflict with some of Hunter's postulates, without however conflicting with the final deduction of Bennett and Hunter—that turnover is the result of self-heating.

The more detailed study made of the thermal properties of point contact diodes has not proved as quantitative as planned, for three reasons. First, analysis becomes very difficult when the whisker is fully allowed for; second the barrier may be non-uniform and ought not therefore to be expected to behave simply and third, the conductivity in the reverse direction may be time dependent (on a scale of microseconds) in a way which is easily confused with dependency on thermal changes. Nonetheless the deductions drawn from the study are not absurd; they yield, for instance, linear dimensions of the barrier of the expected order.

The work suggests new problems, e.g. the comparison of units using the same quality of germanium but a range of dimensions of whisker tip, but none would seem decisive without further consideration of the mechanism of the inverse conduction.

ACKNOWLEDGMENTS

Acknowledgment is made to the Engineer-in-Chief of the General Post Office and to the Controller of Her Majesty's Stationery Office for permission to publish this paper.

REFERENCES

AIGRAIN, P., 1950, *C. R. Acad. Sci., Paris*, **230**, 62.
BENNETT, A. I., and HUNTER, L. P., 1951, *Phys. Rev.*, **81**, 152.
BENZER, S., 1949, *J. Appl. Physics*, **20**, 804.
BILLIG, E., and RIDOUT, M. S., 1951, *Nature, Lond.*, **167**, 1028.
DOUGLAS, R. W., and JAMES, E. G., 1951, *Proc. I. E. E.*, **98**, Part III, 157.
GUNN, J. B., 1952, *Proc. Phys. Soc. B*, **65**, 908.
HENISCH, H. K., and GRANVILLE, J. W., 1951, *Semi-Conducting Materials* (London : Butterworths Scientific Publications), p. 87.
HUNTER, L. P., 1951, *Phys. Rev.*, **81**, 151.
MEACHAM, L. A., and MICHAELS, S. E., 1950, *Phys. Rev.*, **78**, 175.
SIMPSON, J. H., and ARMSTRONG, H. L., 1953, *J. Appl. Phys.*, **24**, 25.

LXXIII. *The Spatial Correlation of Electrons in Atoms and Molecules*
 III: *The Influence of Spin and Antisymmetry on the Correlation of Electrons*

By A. BRICKSTOCK and J. A. POPLE

Department of Theoretical Chemistry, University of Cambridge*

[Received March 20, 1953]

SUMMARY

This paper is concerned with the general restrictions placed on the spatial distribution of electrons by the antisymmetry principle and the specification of spin degeneracy. The total wave function depends on the spatial and spin coordinates of the electrons but, if the quantum-mechanical Hamiltonian is independent of spin, it is only necessary to consider the variation of the wave function in a subspace involving the position coordinates only. The restrictions on this position wave function implied by the condition of antisymmetry and spin degeneracy of the complete wave function are obtained. This analysis involves no approximations. Finally, the theory is used to discuss the description of spatial correlation given by the single-determinant molecular orbital approximate wave function and the way in which this is modified if the interelectronic repulsion is taken into account in a more refined manner.

§ 1. INTRODUCTION

THE antisymmetry principle when applied to any atomic or molecular system, has a profound effect on the correlation or relative distribution of electrons. In its most general form it requires the wave function for an N electron system in the $4N$ -dimensional spin-position configuration space to be antisymmetric for interchange of the coordinates of any pair of electrons. This means, for example, that the value of the wave function for any configuration in which two electrons have the same spin and the same position vectors is zero, so that the probability of such a configuration is also zero. In this way the antisymmetry principle operates to keep electrons of the same spin apart.

Now each electron spin can only have two distinct values, so that the total space may be divided up into 2^N distinct spaces of $3N$ dimensions for specifying the variation of the wave function in terms of the positions of the electrons, given a particular allocation of spins. But although the total wave function is made up of 2^N separate position wave functions, the Hamiltonian is, to a good approximation, independent of spin and most of the physical and chemical properties of the system are, in fact, determined by one of these position wave functions. The problem with which this paper is concerned is to find out the conditions or restrictions on the individual $3N$ -dimensional position wave functions that are implied

* Communicated by Sir J. E. Lennard-Jones, F.R.S.

by the conditions of antisymmetry and spin degeneracy of the complete wave function in $4N$ -dimensions. By eliminating explicit mention of spin wave functions in this way it is possible to simplify the general theory somewhat and to obtain a clearer idea of the effect of the antisymmetry principle on the spatial correlation of electrons.

For a two-electron system the total wave function can be written as the product of a position wave function and a spin wave function. It is then only necessary to note that if the spin function is antisymmetric the position wave function is symmetric and vice versa. This was the method used in discussing two-electron systems in a spherically symmetric field in Parts I and II of this series (Lennard-Jones and Pople 1952, Brickstock and Pople 1952). For systems with more than two electrons however, this is generally no longer possible and it is necessary to develop a more systematic method. In previous theories, this has been done by using orbital approximations for the position wave functions. The molecular orbital theory, for example, is most satisfactorily based on a single determinant of orbitals and spin functions (Lennard-Jones 1949 a, b). This particular approximate wave function has been used by Lennard-Jones (1952) to discuss several examples of electron correlation in atoms and molecules. In the examples considered it is found that, according to the molecular orbital wave function, there is correlation between particles of the same spin but not between electrons of opposite spin. In order to find out how closely these conclusions apply to real systems it is clearly desirable to find the symmetry restrictions on the position wave functions before making approximations, such as the introduction of orbitals.

The method of this paper is based on the condition that if \mathcal{S} is the total spin operator, the total wave function must be an eigen-function of the operators \mathcal{S}^2 and \mathcal{S}_z , \mathcal{S}_z being a component of \mathcal{S} . Given the eigenvalues of \mathcal{S}^2 and \mathcal{S}_z it is possible to obtain explicit equations which must be satisfied by any one of the 2^N position wave functions. Up to this point no approximations are made as long as the Hamiltonian is independent of the spin coordinates. In a later section of the paper the method is used to discuss the general limitations of the description of the correlation of electrons implicit in the single-determinant molecular orbital function.

§ 2. GENERAL PROPERTIES OF THE WAVE FUNCTION FOR A STATE OF GIVEN SPIN DEGENERACY

Following the usual notation, the spin wave function for a single electron j will be written $\alpha(j)$ or $\beta(j)$. Then if \mathcal{S}_j is the spin angular momentum operator in units of $h/2\pi$ and if \mathcal{S}_{jx} , \mathcal{S}_{jy} and \mathcal{S}_{jz} are the corresponding components

$$\left. \begin{array}{ll} 2\mathcal{S}_{jx}\alpha(j)=\beta(j) & 2\mathcal{S}_{jx}\beta(j)=\alpha(j). \\ 2\mathcal{S}_{jy}\alpha(j)=i\beta(j) & 2\mathcal{S}_{jy}\beta(j)=-i\alpha(j). \\ 2\mathcal{S}_{jz}\alpha(j)=\alpha(j) & 2\mathcal{S}_{jz}\beta(j)=-\beta(j). \end{array} \right\} \quad \dots \quad (2.1)$$

The wave function of a many-electron system will be built up from a series of products of the individual spin functions $\alpha(j)$ or $\beta(j)$, each product having a coefficient which is a function of the position coordinates of the electrons. Furthermore if \mathcal{S} is the total spin angular momentum operator

$$\mathcal{S} = \sum_{j=1}^N \mathcal{S}_j \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad (2.2)$$

then the total wave function must be an eigenfunction of the operators \mathcal{S}^2 and \mathcal{S}_z , the corresponding eigenvalues being $S(S+1)$ and S_z where $2S$ is integral and S_z has any of the $(2S+1)$ values $-S, -S+1, \dots, S-1, S$.

For states with a given value of S_z , the only spin functions with non-vanishing coefficients are products of N_1 α -functions and N_2 β -functions where

$$\left. \begin{array}{l} N_1 + N_2 = N, \\ N_1 - N_2 = 2S_z, \end{array} \right\} \quad \dots \quad \dots \quad \dots \quad \dots \quad (2.3)$$

N being the total number of electrons. Suppose we write

$$\Phi(1, 2 \dots N_1 : N_1 + 1, \dots N)$$

for the coefficient of the spin function

$$\alpha(1) \alpha(2) \dots \alpha(N_1) \beta(N_1 + 1) \dots \beta(N)$$

in the total wave function. If only real wave functions are considered, $\Phi(1, 2 \dots N_1 : N_1 + 1, \dots N)$ will be undetermined in sign, but this is unimportant. $[\Phi(1, 2 \dots N_1 : N_1 + 1, \dots N)]^2$ is proportional to the probability of finding electrons with α -spin in volume elements at points $1, 2 \dots N_1$ and electrons with β -spin in volume elements at points $N_1 + 1, N_1 + 2, \dots N$. But since electrons are indistinguishable, this probability is independent of the numbers attached to the electrons. Consequently the coefficient of $\alpha(i)\alpha(j) \dots \alpha(k)\beta(m)\beta(n) \dots \beta(p)$, where $i, j \dots k$ is any N_1 -membered subset of $1, 2 \dots N$ and $m, n \dots p$ is the complementary subset, is $\pm \Phi(i, j \dots k : m, n, \dots p)$, the signs being determined by the antisymmetry principle.

Application of the Antisymmetry Principle

Suppose first that we interchange the coordinates of any two electrons in the group $1 \dots N_1$. This must change the sign of the total wave function and consequently the sign of the coefficient of

$$\alpha(1)\alpha(2) \dots \alpha(N_1)\beta(N_1 + 1) \dots \beta(N).$$

The function $\Phi(1, 2 \dots N_1 : N_1 + 1, \dots N)$, therefore, must be antisymmetric for interchanges of pairs of coordinates *within* either of the two groups. Further, all such functions must be combined into a total wave function which is antisymmetric for any interchange. The correct choice of signs is given by a total wave function

$$\Psi = \sum_P (-1)^P P \Phi(1, 2 \dots N_1 : N_1 + 1, \dots N) \alpha(1) \dots \alpha(N_1) \beta(N_1 + 1) \dots \beta(N), \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad (2.4)$$

where the summation is over all permutations P of the numbers $1, 2 \dots N$, $(-1)^P$ being the parity of the permutation P . Many of the terms in (2.4) will be identical on account of the partial antisymmetry of Φ noted above. There will, in fact, be $N!/N_1! N_2!$ independent terms corresponding to the number of independent spin wave functions with eigenvalue $2S_z = N_1 - N_2$.

Equation (2.4) shows how the total wave function may be obtained from a complete knowledge of one position function

$$\Phi(1, 2 \dots N_1: N_1+1, \dots N).$$

Since the spin factors are orthogonal, it is only necessary to consider one term of (2.4) in working out the matrix elements of any operator (such as the Hamiltonian) which does not involve the spin coordinates.

Specification of Spin Degeneracy

An important equation to be satisfied by the position wave function $\Phi(1, 2 \dots N_1: N_1+1, \dots N)$ follows from the condition that Ψ (eqn. (2.4)) must be an eigenfunction of \mathcal{S}^2 corresponding to the eigenvalue $S(S+1)$, that is

$$\mathcal{S}^2\Psi = S(S+1)\Psi. \dots \dots \dots \quad (2.5)$$

Using

$$\mathcal{S}^2 = \sum_i \mathcal{S}^2 + 2 \sum_{i < j} \{ \mathcal{S}_x \mathcal{S}_{jx} + \mathcal{S}_{iy} \mathcal{S}_{jy} + \mathcal{S}_z \mathcal{S}_{jz} \}, \dots \dots \quad (2.6)$$

and eqns. (2.1), (2.5) become

$$\begin{aligned} \sum_P (-1)^P P \Phi(1, 2 \dots N_1: N_1+1, \dots N) \Bigg\{ & [N_1(N_1+2) + N_2(N_2+2) \\ & - 2N_1N_2 - 4S(S+1)] \alpha(1) \dots \alpha(N_1) \beta(N_1+1) \dots \beta(N) + 4 \sum_{i=1}^{N_1} \sum_{j=N_1+1}^N \\ & \alpha(1) \dots \alpha(i-1) \alpha(j) \alpha(i+1) \dots \alpha(N_1) \beta(N_1+1) \dots \beta(j-1) \beta(i) \beta(j+1) \\ & \times \dots \beta(N) \Bigg\} = 0. \dots \dots \dots \dots \dots \dots \dots \quad (2.7) \end{aligned}$$

Equating the coefficient of $\alpha(1) \dots \alpha(N_1) \beta(N_1+1) \dots \beta(N)$ to zero we get $[4S(S+1) - N_1(N_1+2) - N_2(N_2+2) + 2N_1N_2]$

$$\begin{aligned} & \times \Phi(1, 2 \dots N_1: N_1+1, \dots N) + 4 \sum_{i=1}^{N_1} \sum_{j=N_1+1}^N \\ & \times \Phi(1, \dots i-1, j, i+1, \dots N_1: N_1+1, \dots j-1, i, j+1, \dots N) = 0. \dots \dots \quad (2.8) \end{aligned}$$

The general result may be summarized by saying that the spin-independent physical properties of a state with spin eigenvalues $S(S+1)$ and S_z may be calculated from a position wave function

$$\Phi(1, 2 \dots N_1: N_1+1, \dots N)$$

which is a solution of the Schrödinger equation, is antisymmetric for interchanges within the groups $1 \dots N$, and $N_1+1, \dots N$ and which satisfies eqn. (2.8).

In the absence of a magnetic field, the $2S+1$ different states corresponding to a given value of S will be degenerate and we may, without loss of generality, only consider one of them. It is convenient to take the state with $S_z=+S$ corresponding to as many electrons with α -spin as possible. Then

$$\left. \begin{array}{l} N_1 = \frac{1}{2}N + S, \\ N_2 = \frac{1}{2}N - S, \end{array} \right\} \quad \dots \dots \dots \quad (2.9)$$

and (2.8) becomes

$$\begin{aligned} (S - \frac{1}{2}N)\Phi(1, \dots, \frac{1}{2}N + S : \frac{1}{2}N + S + 1, \dots, N) + \sum_{i=1}^{\frac{1}{2}N+S} \sum_{j=\frac{1}{2}N+S+1}^N \\ \times \Phi(1, \dots, i-1, j, i+1, \dots, \frac{1}{2}N + S : \frac{1}{2}N + S + 1 \dots j-1, i, j+1, \dots, N) \\ = 0. \quad \dots \dots \dots \quad (2.10) \end{aligned}$$

§ 3. APPROXIMATE DETERMINANTAL WAVE FUNCTIONS AND THE CORRELATION OF ELECTRONS

The most commonly used approximate wave function for the ground states of molecules is the single determinant molecular orbital function. These states are usually singlets, the number of electrons being even. The total wave function Ψ is built up from $\frac{1}{2}N$ orbitals (one-electron functions) $\phi_1 \dots \phi_{\frac{1}{2}N}$ in combination with α or β spin functions, the whole being arranged in a determinant to satisfy the antisymmetry principle. This can be written in the form

$$\begin{aligned} \Psi_{\text{det}} = \det \{ \phi_1(1)\alpha(1)\phi_2(2)\alpha(2) \dots \phi_{\frac{1}{2}N}(\frac{1}{2}N)\alpha(\frac{1}{2}N)\phi_1(\frac{1}{2}N+1) \\ \times \beta(\frac{1}{2}N+1) \dots \phi_{\frac{1}{2}N}(N)\beta(N) \}. \quad \dots \dots \dots \quad (3.1) \end{aligned}$$

Single determinants can also be used to represent some excited states. The single determinant function for certain states with $S_z=+S$ (that is with as many α -spins as possible) can be written

$$\begin{aligned} \Psi_{\text{det}} = \det \{ \phi_1(1)\alpha(1) \dots \phi_{N_1}(N_1)\alpha(N_1)\phi_1(N_1+1)\beta(N_1+1) \dots \\ \times \phi_{N_2}(N_1+N_2)\beta(N_1+N_2) \}, \quad \dots \dots \dots \quad (3.2) \end{aligned}$$

N_1 and N_2 being given by (2.9).

If we expand this determinant, the corresponding approximation to the position wave function Φ discussed in § 2 is easily seen to be

$$\begin{aligned} \Phi_{\text{det}}(1, 2 \dots N_1 : N_1+1, \dots N_1+N_2) = \det \{ \phi_1(1)\phi_2(2) \dots \phi_{N_1}(N_1) \} \\ \times \det \{ \phi_1(N_1+1) \dots \phi_{N_2}(N_1+N_2) \}. \quad \dots \dots \dots \quad (3.3) \end{aligned}$$

This wave function is clearly antisymmetric for interchanges within the groups $1, 2 \dots N_1$ and $N_1+1, \dots N_1+N_2$. To confirm that (3.2) is an eigenfunction of \mathcal{S}^2 with the eigenvalue $S(S+1)$ it is necessary to show that (3.3) satisfies the general eqn. (2.8). This is easily demonstrated if the determinants are expanded.

We have already noted that the single position wave function $\Phi(1, 2 \dots N_1 : N_1+1, \dots N_1+N_2)$ gives the same description of the spatial distribution of electrons as the complete function Ψ . Now since

electrons $1 \dots N_1$ are associated with α -spin and $N_1+1, \dots N_1+N_2$ with β -spin, it is immediately clear from (3.3) that, as the wave function separates into a simple product, there is no correlation between electrons of opposite spin, according to the single determinant function. This result was noted in several examples by Lennard-Jones (1952), but from this analysis it is seen to apply quite generally.

Up to this point, no explicit mention has been made of the form of interaction between the electrons except to postulate that it is independent of spin. In fact electrons repel each other according to a Coulomb law and this will tend to keep them apart whatever their spin. The failure of the single determinant function to allow for this correlation between electrons of opposite spin is one of its major disadvantages and it is a problem of considerable importance to find the magnitude of such a correlation and the effect it has on the properties of the system.

One method of investigating the limitations of the determinantal wave function (3.3) is to consider functions of the type

$$\begin{aligned} \Phi(1, 2 \dots N_1 : N_1+1, \dots N_1+N_2) \\ = \Phi_{\text{det}}(1, 2 \dots N_1 : N_1+1, \dots N_1+N_2) Q(1, 2 \dots N_1 : N_1+1, \dots \\ N_1+N_2), \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad (3.4) \end{aligned}$$

where $Q(1, 2 \dots N_1 : N_1+1, \dots N_1+N_2)$ is some correcting function containing parameters whose optimum values can be estimated by the variation method. Since $\Phi(1, 2 \dots N_1 : N_1+1, \dots N_1+N_2)$ must be antisymmetric with respect to interchanges within the groups $1, 2 \dots N_1$ and $N_1+1, \dots N_1+N_2$, Q must be symmetric with respect to such interchanges. It remains to find what limitations are imposed on Q by the general condition (2.8). Substituting (3.4) into (2.8) and using the fact that Φ_{det} satisfies (2.8) we obtain

$$\begin{aligned} \sum_{i=1}^{N_1} \sum_{j=N_1+1}^{N_1+N_2} \Phi_{\text{det}}(1, \dots i-1, j, i+1, \dots N_1 : N_1+1, \dots j-1, i, j+1, \dots N) \\ \times \{Q(1, \dots i-1, j, i+1, \dots N_1 : N_1+1, \dots j-1, i, j+1 \dots N) \\ - Q(1, 2 \dots N_1 : N_1+1, \dots N)\} = 0. \quad \dots \quad \dots \quad \dots \quad (3.5) \end{aligned}$$

A sufficient condition for this equation to be satisfied identically is that $Q(1, 2 \dots N_1 : N_1+1, \dots N)$ is symmetric with respect to all interchanges including those between the two groups. The authors have not been able to prove that this condition is necessary in the general case, although it is so for a wide class of functions Q . If we write

$$\begin{aligned} Q(1, 2 \dots N_1 : N_1+1, \dots N) \\ = 1 + \lambda_1 \sum_{i < j=1}^{N_1 N_2} f(i, j) + \lambda_2 \sum_{i < j=N_1+1}^{N_1 N_2} f(i, j) + \mu \sum_{i=1}^{N_1} \sum_{j=N_1+1}^N f(i, j), \quad \dots \quad (3.6) \end{aligned}$$

where $f(i, j)$ is some symmetric function of coordinates i and j only, then in general it is necessary that $\lambda_1 = \lambda_2 = \mu$ if (3.5) is to be satisfied. This is proved in the appendix.

If the function $f(i, j)$ is chosen so as to increase in value as electrons i and j move away from one another (if, for example, $f(i, j)$ is equal to the

interelectronic distance r_{ij}), the correcting function Q should improve the total wave function by allowing for the tendency of electrons to keep apart because of their electrostatic repulsion. The first two sums in (3.6) can be interpreted as modifying the correlation between electrons of the same spin, while the third introduces correlation between electrons of opposite spin. The equality of the three coefficients means, therefore, that if the determinantal wave function is corrected to allow for the repulsion between electrons of opposite spin, then the tendency for electrons of the same spin to keep apart will be further increased. In other words there will be a sharpening of the pattern of electron distribution predicted by the single determinant function. This general result will be illustrated for the simple systems of a closed shell of electrons restricted to a sphere in the next paper.

APPENDIX

We give here the proof that if $Q(1, 2 \dots N_1; N_1+1, \dots N)$ has the particular form (3.6), then in general it follows that $\lambda_1 = \lambda_2 = \mu$. (3.5) can be written

$$\sum_{l < m}^N c_{lm} f(l, m) = 0, \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad (A.1)$$

where c_{lm} involve the functions within the determinants but not the $f(i, j)$. Now if the $f(i, j)$ are linearly independent, then, in general, it follows from (A.1) that

for all l, m . (If there are certain functional relationships between the ϕ_p and $f(l, m)$, (A.2) may not be a necessary consequence of (A.1), but these cases would be exceptional.) Using (3.6) we find

$$Q(1 \dots i-1, j, i+1, \dots N_1: N_1+1, \dots j-1, i, j+1, \dots N) \\ - Q(1, \dots N_1: N_1+1, \dots N) = (\lambda_1 - \mu) \sum_{r=1 (r \neq i)}^{N_1} [f(j, r) - f(i, r)] \\ + (\lambda_2 - \mu) \sum_{s=N_1+1 (s \neq j)}^N [f(i, s) - f(j, s)]. \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad (A.3)$$

If l and m are both less than N_1 , the explicit expression for c_{lm} is

$$c_{lm} = (\mu - \lambda_1) \left\{ \sum_{j=N_1+1}^N [\Phi_{\text{det}}(1, \dots, l-1, j, l+1, \dots, N_1; N_1+1, \dots, j-1, l, j+1, \dots, N) \right. \\ \left. + \Phi_{\text{det}}(1, \dots, m-1, j, m+1, \dots, N_1; N_1+1, \dots, j-1, m, j+1, \dots, N)] \right\} \quad (\text{A.4})$$

whence

$$\sum_{l=1}^{N_1} \sum_{m=1}^{N_1} c_{lm} = 2N_1(\mu - \lambda_1) \sum_{i=1}^{N_1} \sum_{j=N_1+1}^{N_1} \times \Phi_{\text{det}}(1, \dots, i-1, j, i+1, \dots, N_1 : N_1+1, \dots, j-1, i, j+1, \dots, N). \quad (\text{A.5})$$

Also since Φ_{det} satisfies condition (2.10) itself, this can be written

$$\sum_{l=1}^{N_1} \sum_{m=1}^{N_1} c_{lm} = 2N_1 N_2 (\mu - \lambda_1) \Phi_{\text{det}}(1, \dots, N_1; N_1 + 1, \dots, N). \quad \dots \quad (\text{A.6})$$

But this expression must vanish by (A.2) so that $\mu = \lambda_1$. Similarly $\mu = \lambda_2$.

REFERENCES

BRICKSTOCK, A., and POPLE, J. A., 1952, *Phil. Mag.*, **43**, 1090 (Part II).
 LENNARD-JONES, Sir J., 1949 a, *Proc. Roy. Soc.*, **198**, 1; 1949 b, *Ibid.*, **198**, 14; 1952, *J. Chem. Phys.*, **20**, 1024.
 LENNARD-JONES, Sir J., and POPLE, J. A., 1952, *Phil. Mag.*, **43**, 581 (Part I).

LXXIV. *The Spatial Correlation of Electrons in Atoms and Molecules*IV: *The Correlation of Electrons on a Spherical Surface*

By A. BRICKSTOCK and J. A. POPLE

Department of Theoretical Chemistry, University of Cambridge*

[Received March 20, 1953]

SUMMARY

The aim of this paper is to compare the effects of the exclusion principle and the interelectronic repulsion on the correlation of electrons when they are restricted to move on a spherical surface. The two examples considered are four electrons of the same spin and eight paired electrons, four of each spin. These are analogous to the 5S state of carbon and the neon closed shell respectively. Lennard-Jones (1952), using a single determinant wave function has shown that the correlation of electrons in these systems is such that particles with the same spin tend to arrange themselves in a tetrahedral configuration relative to one another, but that in the closed shell there is no correlation between the two tetrahedra. In this paper a more accurate wave function is used, making some allowance for the correlation of electrons due to electrostatic repulsion. It is found that the probability of electrons of the same spin having a tetrahedral distribution is increased while at the same time there is a tendency for electrons of opposite spin to keep apart. The magnitude of this correlation between positions of electrons with opposite spin, however, is much smaller than that between those of the same spin due directly to the exclusion principle.

§ 1. INTRODUCTION

THE relation between the spatial correlation of electrons in atoms and the nature of directed valence has been the subject of several recent investigations. It has become clear that the exclusion principle leading to the antisymmetry condition on the wave function has an important direct effect on the distribution of electrons relative to each other. The simplest wave function satisfying the antisymmetry condition is formed from a single determinant of orbitals associated with either α or β spin. The type of correlation implied by a function of this form has been discussed in general and for several examples by Lennard-Jones (1949, 1952). It is found that the effect of the antisymmetry condition is to

* Communicated by Sir J. E. Lennard-Jones, F.R.S.

separate electrons of the same spin so that, at any given instant, they will tend to be in different parts of a molecule (in different bonds, for example). According to this approximation, however, electrons of different spin still move independently. Examples discussed by Lennard-Jones (1952) include the quintuplet state of carbon and neon-like closed shells. In the 5S carbon state, the four outer electrons have the same spin and, according to the single determinant approximation, keep apart from one another so that they are most likely to be found in a tetrahedral configuration. In a neon-like closed shell there are four electrons of each spin and again according to the single determinant function, each group of four tend to keep to a tetrahedral configuration. On the other hand, there is no correlation between the positions of the two tetrahedra.

The principal aim of this paper is to investigate how these conclusions are modified if more accurate wave functions are used and, in particular, to find the extent of correlation between electrons of opposite spin. In order to make it feasible to carry out calculations beyond the single-determinant approximation, the problem is simplified to the motion of four or eight electrons on the surface of a sphere. Although the numerical results are not directly applicable to the 5S state of carbon or the neon closed shell, they do give a qualitative indication of the sort of modifications of the angular distribution that are to be expected.

§ 2. THE MOTION OF FOUR ELECTRONS ON A SPHERICAL SURFACE

The Hamiltonian of a system of n electrons on a spherical surface of radius a can be written

$$\mathcal{H} = -\frac{1}{2} \sum_{i=1}^n A_i + \sum_{i < j}^n (1/r_{ij}), \quad \dots \quad (2.1)$$

where A_i is the operator

$$A_i = \frac{1}{a^2 \sin \theta_i} \frac{\partial}{\partial \theta_i} \left(\sin \theta_i \frac{\partial}{\partial \theta_i} \right) + \frac{1}{a^2 \sin^2 \theta_i} \frac{\partial^2}{\partial \phi_i^2}, \quad \dots \quad (2.2)$$

where (a, θ_i, ϕ_i) are spherical polar coordinates of electron i . It is unnecessary to include a term due to the attraction of a central nucleus as this would be a constant. The interelectronic repulsion energy $(1/r_{ij})$ can be expanded in the form

$$1/r_{ij} = a^{-1} \sum_{l=0}^{\infty} P_l(\cos \theta_{ij}), \quad \dots \quad (2.3)$$

where θ_{ij} is the angle subtended by i and j at the centre and $P_l(\cos \theta)$ is a Legendre polynomial of order l , normalized so that $P_l(1) = 1$.

For an S-state in which all four electrons have the same spin (α say), the total wave function can be written as a product of a spatial wave function Φ and a spin function $\alpha(1)\alpha(2)\alpha(3)\alpha(4)$. Then Φ must be anti-symmetric in the spatial coordinates of the four electrons. The single

determinant approximation Φ_0 (corresponding to the configuration sp^3) can be written in the form

$$\Phi_0 = \frac{3}{2\sqrt{2}} \frac{1}{a^5} \frac{1}{(4\pi)^2} \begin{vmatrix} 1 & x_1 & y_1 & z_1 \\ 1 & x_2 & y_2 & z_2 \\ 1 & x_3 & y_3 & z_3 \\ 1 & x_4 & y_4 & z_4 \end{vmatrix} \quad \dots \quad (2.4)$$

where (x_i, y_i, z_i) are Cartesian coordinates of electron i referred to the centre of the sphere as origin. The wave function (2.4) is, of course, independent of direction of these axes. Since the particles all lie on the sphere of radius a

$$x_i^2 + y_i^2 + z_i^2 = a^2. \quad \dots \quad (2.5)$$

The error implicit in the single-determinant wave function is measured by the ratio Φ/Φ_0 . This can be expanded in a general power series

$$\begin{aligned} \Phi/\Phi_0 = 1 + & (1\text{st degree polynomial in } x_i, y_i, z_i) + \\ & (2\text{nd degree polynomial in } x_i, y_i, z_i) + \dots \quad (2.6) \end{aligned}$$

Now all polynomials of odd degree in this expression vanish identically because terms like $x_i\Phi_0$ and Φ_0 behave differently under inversion. Similarly terms such as x_iy_j do not appear in the second degree polynomial since $x_iy_j\Phi_0$ and Φ_0 have different symmetry properties with respect to reflection in the plane $x=0$. Further, all coefficients of x_ix_j , y_iy_j and z_iz_j must be equal since there are no preferred axes. Terms of the type x_i^2 would only occur in the combination $(x_i^2 + y_i^2 + z_i^2)$ which is a constant and can be included in the term of zero degree. Noting that

$$\cos \theta_{ij} = a^{-2}(x_i x_j + y_i y_j + z_i z_j), \quad \dots \quad (2.7)$$

it becomes possible to expand Φ/Φ_0 in the form

$$\Phi/\Phi_0 = 1 + \lambda \sum_{i < j} \cos \theta_{ij} + (4\text{th degree terms}). \quad \dots \quad (2.8)$$

The simplest method of improving on the single determinant approximation, therefore, is to neglect the fourth degree terms and use a trial wave function of the form

$$\Phi = \Phi_0 + \lambda \Phi_1,$$

$$\Phi_1 = \Phi_0 \sum_{i < j} \cos \theta_{ij}, \quad \dots \quad (2.9)$$

the optimum value of λ being determined by minimizing the energy

$$\int \Phi \mathcal{H} \Phi d\tau / \int \Phi^2 d\tau.$$

The matrix elements of the Hamiltonian (H_{ij}) and unity (S_{ij}) are found to be

$$\left. \begin{array}{ll} H_{00} = 3a^{-2} + \frac{22}{5}a^{-1} & S_{00} = 1 \\ H_{01} = -3a^{-2} - \frac{142}{35}a^{-1} & S_{01} = -1 \\ H_{11} = \frac{41}{5}a^{-2} + \frac{1364}{175}a^{-1} & S_{11} = \frac{8}{5} \end{array} \right\} \quad \dots \quad (2.10)$$

To illustrate the results a value of 1.2 atomic units was taken for a (approximately the most probable radius of the outer electrons in the carbon atom). The optimum value of λ was found to be -0.08685 .

§3. THE MOTION OF EIGHT PAIRED ELECTRONS ON A SPHERICAL SURFACE

The ground state of a system of eight electrons on a sphere will be 1S . As all the electrons are paired, the only spin functions occurring in the total wave function are those with an equal number of α and β types such as $\alpha(1)\alpha(2)\alpha(3)\alpha(4)\beta(5)\beta(6)\beta(7)\beta(8)$. As all such spin functions are orthogonal, it is only necessary to consider the position wave function associated with one such as

$$\Phi(1, 2, 3, 4 : 5, 6, 7, 8) \alpha(1)\alpha(2)\alpha(3)\alpha(4)\beta(5)\beta(6)\beta(7)\beta(8). \quad \dots \quad (3.1)$$

The general symmetry properties of $\Phi(1, 2, 3, 4 : 5, 6, 7, 8)$ have been discussed in an earlier part of this series (Brickstock and Pople 1953).

The single determinant approximation to the total wave function (corresponding to the configuration s^2p^6) gives the following approximate form (Φ_0) for Φ

$$\Phi_0 = \frac{9}{8} \frac{1}{a^{10}} \frac{1}{(4\pi)^4} \left| \begin{array}{cccc} 1 & x_1 & y_1 & z_1 \\ 1 & x_2 & y_2 & z_2 \\ 1 & x_3 & y_3 & z_3 \\ 1 & x_4 & y_4 & z_4 \end{array} \right| \times \left| \begin{array}{cccc} 1 & x_5 & y_5 & z_5 \\ 1 & x_6 & y_6 & z_6 \\ 1 & x_7 & y_7 & z_7 \\ 1 & x_8 & y_8 & z_8 \end{array} \right| \quad \dots \quad (3.2)$$

This wave function is rather simpler than the complete 8×8 determinant of the total wave function and gives the same description of the correlative motion of the electrons. As Φ_0 is a simple product of two determinants, each involving coordinates of electrons of a given type of spin, it is immediately clear that this function leads to no correlation between electrons of opposite spin as found by Lennard-Jones (1952).

Following the same method as was used in § 2 for four electrons, an improved description is obtained by using a trial wave function

$$\left. \begin{array}{l} \Phi = \Phi_0 + \lambda \Phi_1, \\ \Phi_1 = \Phi_0 \sum_{i < j} \cos \theta_{ij}, \end{array} \right\} \quad \dots \quad (3.3)$$

λ being chosen so as to minimize the energy. It can be shown (Brickstock and Pople 1952) that if Φ/Φ_0 is expanded in the form

$$\Phi/\Phi_0 = 1 + \lambda \left\{ \sum_{i < j=1}^4 \cos \theta_{ij} + \sum_{i < j=5}^8 \cos \theta_{ij} \right\} + \mu \sum_{i=1}^4 \sum_{j=5}^8 \cos \theta_{ij} + (4\text{th degree terms in } x_i, y_i, z_i) + \dots \quad \dots \quad \dots \quad \dots \quad (3.4)$$

then λ must be equal to μ if the total wave function is to be that of a singlet state. The trial function (3.3), therefore, represents the best function obtainable if the terms of fourth and higher degree in (3.4) are neglected. The matrix elements H_{ij} and S_{ij} are found to be

$$\left. \begin{array}{ll} H_{00} = 6a^{-2} + \frac{124}{5}a^{-1} & S_{00} = 1 \\ H_{01} = -12a^{-2} - \frac{4996}{105}a^{-1} & S_{01} = -2 \\ H_{11} = \frac{152}{3}a^{-2} + \frac{422704}{2625}a^{-1} & S_{11} = \frac{98}{15} \end{array} \right\} \quad \dots \quad \dots \quad \dots \quad (3.5)$$

To illustrate the results the secular equation

$$|H_{ij} - E S_{ij}| = 0 \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad (3.6)$$

has been solved for $a = 0.7$ atomic units, corresponding approximately to the most probable radius of the outer electrons in neon. This leads to $\lambda = -0.10177$.

§ 4. THE CORRELATION OF ELECTRONS ON A SPHERICAL SURFACE

As pointed out in § 1, the relative position of four electrons of the same spin on a spherical surface is closely related to the correlation of the outer electrons of the carbon atom in its 5S state. The type of correlation predicted by a single-determinant wave function for this system has been discussed by Lennard-Jones (1952). Using approximate Slater orbitals of the form

$$f(r), f(r)\sqrt{3} \cos \theta, f(r)\sqrt{3} \sin \theta \cos \phi, f(r)\sqrt{3} \sin \theta \sin \phi \quad \dots \quad (4.1)$$

he shows that the probability of finding an electron in a volume element dv_2 at (r_2, θ_2, ϕ_2) and of simultaneously finding another in a volume element dv_1 at (r_1, θ_1, ϕ_1) is proportional to

$$P^{\alpha\alpha}(1, 2) = \{f(r_1)f(r_2)\}^2 \{16 - (1 + 3 \cos \theta_{12})^2\}, \quad \dots \quad \dots \quad \dots \quad (4.2)$$

where θ_{12} is the angle subtended at the centre. If the particles are restricted to a sphere, the angular dependence of (4.2) is unaltered and the probability of finding one electron in a solid angle $d\omega_1$ and another in $d\omega_2$ is $P^{\alpha\alpha}(1, 2) d\omega_1 d\omega_2$ where

$$P^{\alpha\alpha}(1, 2) = \{16 - (1 + 3 \cos \theta_{12})^2\} / 16\pi^2. \quad \dots \quad \dots \quad \dots \quad (4.3)$$

For the eight-electron problem, in which the orbitals are doubly occupied, this function gives the probability of finding one electron with α -spin in $d\omega_1$ and another of α -spin in $d\omega_2$ simultaneously. There is a similar function $P^{\alpha\beta}$ for electrons of β -spin. The probability of finding an electron of α -spin in $d\omega_1$ and another of β -spin in $d\omega_2$ is $P^{\alpha\beta}(1, 2) d\omega_1 d\omega_2$. According to the single determinant wave function

$$P^{\alpha\beta}(1, 2) = \pi^{-2} \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \quad (4.4)$$

We shall now consider how (4.3) and (4.4) are modified by improvement of the wave function. It is easily shown that

$$P^{\alpha\alpha}(1, 2) = 12 \int \Phi^2 d\tau_3 d\tau_4 / \int \Phi^2 d\tau_1 \dots d\tau_4$$

or $12 \int \Phi^2 d\tau_3 \dots d\tau_8 / \int \Phi^2 d\tau_1 \dots d\tau_8 \quad \cdot \cdot \cdot \quad (4.5)$

in the four and eight electron cases respectively and that

$$P^{\alpha\beta}(1, 5) = 16 \int \Phi^2 d\tau_2 d\tau_3 d\tau_4 d\tau_6 d\tau_7 d\tau_8 / \int \Phi^2 d\tau_1 \dots d\tau_8. \quad \cdot \cdot \cdot \quad (4.6)$$

The Four-Electron Problem

When expressed in terms of Legendre polynomials, (4.3) can be written

$$\frac{4\pi^2}{3} P^{\alpha\alpha}(1, 2) = 1 - \frac{1}{2} P_1(\cos \theta_{12}) - \frac{1}{5} P_2(\cos \theta_{12}). \quad \cdot \cdot \cdot \quad (4.7)$$

This function is zero for $\theta_{12}=0$ and has a maximum at the tetrahedral angle $\cos^{-1}(-\frac{1}{3})$, indicating that the tetrahedral configuration is the most probable. On introducing ϕ_1 with $a=1.2$ atomic units, we find

$$\frac{4\pi^2}{3} P^{\alpha\alpha}(1, 2) = 1 - 0.5366 P_1(\cos \theta_{12}) - 0.5091 P_2(\cos \theta_{12})$$

$$+ 0.0429 P_3(\cos \theta_{12}) + 0.0028 P_4(\cos \theta_{12}). \quad \cdot \cdot \cdot \quad (4.8)$$

This function is again zero if $\theta_{12}=0$ and has its maximum near the tetrahedral angle. However, the actual value of (4.8) at $\cos \theta_{12}=-\frac{1}{3}$ is increased from 1.3333 to 1.3627. This indicates that the probability function is 'sharpened' in the neighbourhood of the tetrahedral angle, the tetrahedral configuration being made more probable at the expense of other configurations.

The Eight-Electron Problem

Using the results for $a=0.7$ quoted in the previous section, it is found that

$$\frac{4\pi^2}{3} P^{\alpha\alpha}(1, 2) = 1 - 0.5347 P_1(\cos \theta_{12}) - 0.5100 P_2(\cos \theta_{12})$$

$$+ 0.0417 P_3(\cos \theta_{12}) + 0.0030 P_4(\cos \theta_{12}), \quad \cdot \quad (4.9)$$

and $\pi^2 P^{\alpha\beta}(1, 5) = 1 - 0.0375 P_1(\cos \theta_{15}) + 0.0002 P_2(\cos \theta_{15}). \quad \cdot \quad (4.10)$

Equation (4.9) is similar to the corresponding result for the four-electron problem, so that similar conclusions apply, the probability of the tetrahedral distribution being increased by the improvement of the wave function. The correlation function for electrons of different spin (4.10) is no longer constant, but slightly favours negative values of $\cos \theta_{15}$. This means that there is a tendency for a given pair of electrons of opposite spin to be found on opposite sides of the nucleus. The magnitude of the effect, however, is small, the average value of θ_{15} as calculated from (4.10) being only 90.8° .

These calculations on the simplified spherical surface model give general support to the conclusions reached by Lennard-Jones (1952) about the angular distribution of electrons in the 5S state of carbon and closed shells of electrons as in the inert gases. It is clear that the correlation of electrons due to the exclusion principle, leading to the tetrahedral configuration of electrons of the same spin, is more important than the effect of electrostatic repulsion between electrons. According to the theory of this paper the electrons of each spin in an inert gas closed shell may be described as tending to take up a tetrahedral configuration among themselves, there being only a small correlation between the positions of the two tetrahedra.

The authors are indebted to Sir John Lennard-Jones for suggesting this work.

REFERENCES

BRICKSTOCK, A., and POPPLE, J. A., 1953, *Phil. Mag.*, **44**, 697 (Part III).
LENNARD-JONES, Sir J., 1949, *Proc. Roy. Soc. A*, **198**, 14; 1952, *J. Chem. Phys.*, **20**, 1024.

LXXV. *Collisional Effects and the Conduction Current in an Ionized Gas*

By K. C. WESTFOLD
University of Sydney*

[Received March 21, 1953]

ABSTRACT

A first approximation to the transport equation for the conduction current in a binary ionized gas is derived by the iterative methods of Enskog and Chapman. The collision-integrals that arise are evaluated in terms of the current density, as suggested by the results of 'free-path' theory. It is found that the collisional damping factor is a weighted mean of the electron and ion collision frequencies instead of the electron collision frequency. In practice the difference amounts to a factor $4/3$, as predicted in an earlier paper. When applied to static fields the equation yields Chapman and Cowling's formulae without further calculation.

Corresponding approximations are made to the equations of conservation, motion and thermal energy. With Maxwell's equations these provide a reliable set for the investigation of interactions between an ionized gas in motion and the associated radiation field.

The results are applicable to the solar atmosphere and the H II regions of interstellar space. With slight modifications they are also made applicable to a slightly ionized gas such as the H I regions and the lower ionosphere.

§ 1. INTRODUCTION

IT is now generally agreed (see Bhatnagar, Krook and Menzel 1952, p. 35) that the 'steady' component of the radio-frequency radiation received from the Sun is adequately accounted for by the thermal process of free-free electron-ion collisions in the ionized solar atmosphere. The theory of the 'non-thermal' component, and of radiation from the 'radio stars', is in a less satisfactory state.

Perhaps the most attractive hypothesis considered so far, due to Shklovsky (1946) and Martyn (1947), is that this radiation has its origin in macroscopic plasma oscillations within the solar atmosphere. Unfortunately, no satisfactory mechanism of escape of radiation of the plasma frequency from the region of origin has yet been propounded. However, if the oscillations have a range of frequencies extending above the local value of the plasma frequency, as in the case of the transients

* Communicated by Professor S. Chapman, F.R.S.

considered by Jaeger and Westfold (1949), part of the radiation generated can escape along trajectories that pass through the region of origin (Burkhardt and Schlüter 1949, Jaeger and Westfold 1950). The same applies to any other process that produces radiation with an appreciable part of its frequency-spectrum above the local plasma value.

It has been pointed out (Westfold 1951 a) that any macroscopic motion of an ionized gas will be coupled to a non-thermal radiation field with a determinate spectrum. This coupling is exhibited by the equation of motion of the gas and a generalized Lorentz equation for the conduction current. Both of these are non-linear in the mass velocity and the conduction current density, and the latter equation also involves undetermined collision-integrals. In addition, to specify the state of the gas and the associated radiation field, Maxwell's equations and the equations of continuity, conservation of charge and thermal energy are required.

The resulting set of equations appears to be mathematically intractable. In order to make progress, previous investigators (e.g. Bailey 1948, Bohm and Gross 1949) have neglected non-linear terms and either assigned to the pressure tensors and the collision-integrals values suggested by 'free-path' kinetic theory, or neglected them. Although this procedure has the merit of simplifying the equations, its justification and physical significance are by no means clear.

The aim of the present paper is to obtain an adequate simplification of the set by following the well-defined Enskog-Chapman scheme, as presented in the standard monograph* by Chapman and Cowling (1939). Successive approximations to the solutions of Boltzmann's equations for a binary ionized gas are substituted into the corresponding transport equations, enabling consistent first approximations to the constitutive equations for the medium under gravitational, electric and magnetic fields to be obtained. This procedure amounts to the investigation of the first-order effects of departures from the Maxwellian form towards which the distributions of the constituent gases naturally tend. Taken with the equations of the electromagnetic field, these transport equations provide a reliable set for investigating the interactions between an ionized gas in motion and the associated radiation field. The electro-magneto-ionic magneto-ionic and Lorentz theories emerge as special cases.

Although the present paper is particularly concerned with the presence of a radiation field, its results are obtained in a form that is immediately applicable to the case of stationary currents in static electric and magnetic fields, considered in C. & C., ch. 18 and by Cowling (1945). It provides an extension of some of the methods and results of C. & C., ch. 18, which avoids the necessity for separate investigations, *ad hoc* to each case.

In the interests of terseness, the notation of C. & C. is followed and many of their results are quoted direct.

*This book is subsequently referred to as C. & C.

§ 2. THE EQUATIONS OF CONSERVATION, MOTION AND THERMAL ENERGY, AND THE GENERALIZED LORENTZ EQUATION

In kinetic theory, a transport equation for the rate of change of the average value of some molecular property ϕ_s of the s -th constituent gas, is derived from Boltzmann's equation by multiplying by ϕ_s and integrating over the velocity range. There is difficulty about formulating Boltzmann's equations for an ionized gas, because the long-range Coulomb interaction forces between the molecules make it questionable to reckon a collision as wholly binary (see C. & C., p. 178, and Bhatnagar, Krook and Menzel 1952, p. 4). The difficulty is met by limiting the range of the collision parameter b , the distance from the centre to an asymptote of the relative orbit of two interacting molecules, in a somewhat arbitrary manner. However, it happens that in applications the values of the desired physical quantities are not very sensitive to the values of the end points of the range. The effects of multiple interactions beyond this range of b are smoothed out and accounted for by the macroscopic electric vector which, with any external fields imposed on the medium, determines the motion of molecules between collisions.

It has been shown (Cohen, Spitzer and Routly 1950) that the aggregate effect of a number of small-deflection encounters suffered by a molecule can be appreciable in this connection. It may later prove necessary to modify the analysis of this paper by taking this effect into account.

We restrict our consideration to a binary ionized gas in a gravitational field \mathbf{F} , and electric field \mathbf{E} and a magnetic field \mathbf{H} . \mathbf{E} may contain the electric vector of an electromagnetic field whose magnetic vector is, by comparison, negligibly small. The investigation could be extended to a multiple gas following Cowling (1945), but the present results seem to be adequate for most purposes.

Then Boltzmann's equation for the velocity-distribution function $f_1(\mathbf{C}_1, \mathbf{r}, t)$ is (C. & C., p. 329) :

$$\frac{Df_1}{Dt} + \mathbf{C}_1 \cdot \frac{\partial f_1}{\partial \mathbf{r}} + \left\{ \mathbf{F} + \frac{e_1}{m_1} (\mathbf{E} + \mathbf{c}_0 \wedge \mathbf{H}) - \frac{D\mathbf{c}_0}{Dt} \right\} \cdot \frac{\partial f_1}{\partial \mathbf{C}_1} + \frac{e_1}{m_1} \mathbf{C}_1 \wedge \mathbf{H} \cdot \frac{\partial f_1}{\partial \mathbf{C}_1} - \frac{\partial f_1}{\partial \mathbf{C}_1} \mathbf{C}_1 : \frac{\partial}{\partial \mathbf{r}} \mathbf{c}_0 = -J_1(f_1 f) - J_{12}(f_1 f_2), \quad \dots \quad (2.1)$$

where \mathbf{C}_1 is the peculiar velocity of a molecule of the first constituent and \mathbf{c}_0 the mass velocity; a similar equation governs f_2 . The charges e_1, e_2 are in electromagnetic units.

Substitution of the summational invariants 1, $m_1 \overline{\mathbf{C}_1}$, $\frac{1}{2} m_1 \overline{\mathbf{C}_1^2}$, etc., in the corresponding equations of change yield the equations (C. & C., p. 323) expressing the conservation of the numbers of each constituent, the equation of mass motion of the gas and the equation of thermal energy. To obtain the transport equation for the partial conduction current

$$\mathbf{j}_1 = n_1 e_1 \overline{\mathbf{C}_1}, \quad \dots \quad (2.2)$$

we take

$$\boldsymbol{\varphi}_1 = e_1 \mathbf{C}_1 \quad \dots \quad (2.3)$$

in the equation of change and obtain a result (Westfold 1951 a) which, after substitution from the equation of motion, becomes

$$\frac{D\mathbf{j}_1}{Dt} + \mathbf{j}_1 \frac{\partial}{\partial \mathbf{r}} \cdot \mathbf{c}_0 + \mathbf{j}_1 \cdot \frac{\partial}{\partial \mathbf{r}} \mathbf{c}_0 - \frac{e_1 \rho_1 \rho_2}{m_1 \rho} \left\{ \left(\frac{e_1}{m_1} - \frac{e_2}{m_2} \right) (\mathbf{E} + \mathbf{c}_0 \wedge \mathbf{H}) \right. \\ \left. + \left(\frac{\mathbf{j}_1}{\rho_1} - \frac{\mathbf{j}_2}{\rho_2} \right) \wedge \mathbf{H} - \frac{1}{\rho_1} \frac{\partial}{\partial \mathbf{r}} \cdot \mathbf{p}_1 + \frac{1}{\rho_2} \frac{\partial}{\partial \mathbf{r}} \cdot \mathbf{p}_2 \right\} = 4\mathbf{j}_1, \quad \dots \quad (2.4)$$

where the collision integrals $\Delta\mathbf{j}_1$ are given by the formulae of C. & C., p. 66 with (2.3), and $\mathbf{p}_1, \mathbf{p}_2$ are the partial pressure tensors. The generalized Lorentz equation is obtained by adding (2.4) to a similar equation, giving

$$\frac{D\mathbf{j}}{Dt} + \mathbf{j} \frac{\partial}{\partial \mathbf{r}} \cdot \mathbf{c}_0 + \mathbf{j} \cdot \frac{\partial}{\partial \mathbf{r}} \mathbf{c}_0 - \frac{\rho_1 \rho_2}{\rho} \left(\frac{e_1}{m_1 \rho_1} + \frac{e_2}{m_2 \rho_2} \right) \mathbf{j} \wedge \mathbf{H} - \Delta\mathbf{j} \\ = \frac{\rho_1 \rho_2}{\rho} \left(\frac{e_1}{m_1} - \frac{e_2}{m_2} \right) \left\{ \left(\frac{e_1}{m_1} - \frac{e_2}{m_2} \right) (\mathbf{E} + \mathbf{c}_0 \wedge \mathbf{H}) - \frac{1}{\rho_1} \frac{\partial}{\partial \mathbf{r}} \cdot \mathbf{p}_1 + \frac{1}{\rho_2} \frac{\partial}{\partial \mathbf{r}} \cdot \mathbf{p}_2 \right\}, \\ \dots \quad (2.5)$$

$$\text{where } \Delta\mathbf{j} = \Delta\mathbf{j}_1 + \Delta\mathbf{j}_2, \quad \dots \quad (2.6)$$

and we have made use of the results

$$\mathbf{j}_1 = \frac{e_1/m_1}{e_1/m_1 - e_2/m_2} \mathbf{j}, \quad \mathbf{j}_2 = -\frac{e_2/m_2}{e_1/m_1 - e_2/m_2} \mathbf{j} \quad \dots \quad (2.7)$$

for a binary ionized gas.

The eqn. (2.5), with its non-linear terms in \mathbf{c}_0 and \mathbf{j} and undetermined collision-integrals, appears formidable. As outlined in § 1, the usual simplifications amount to the neglect of non-linear terms in this and the equation of motion, the replacement of the pressure tensors by 'hydrostatic' pressures and the adoption of the relation

$$\Delta\mathbf{j} = -\nu\mathbf{j}, \quad \dots \quad (2.8)$$

where ν is the *electron* collision frequency. In the next two sections we replace, and to some extent justify, this speculative procedure by following the Enskog-Chapman method of successive approximations, stopping at the first non-zero approximations to \mathbf{j} and $\Delta\mathbf{j}$.

§ 3. APPROXIMATIONS TO THE TRANSPORT EQUATIONS

The first approximations to the solutions of Boltzmann's equations, $f_1^{(0)}, f_2^{(0)}$, are obtained (C. & C., p. 330) by writing

$$\frac{e_1}{m_1} \mathbf{C}_1 \wedge \mathbf{H} \cdot \frac{\partial f_1^{(0)}}{\partial \mathbf{C}_1} = -J_1(f_1^{(0)} f^{(0)}) - J_{12}(f_1^{(0)} f_2^{(0)}) \quad \dots \quad (3.1)$$

for (2.1), with a similar equation for $f_2^{(0)}$.

The solutions are the Maxwellian functions

$$\left. \begin{aligned} f_1^{(0)} &= n_1 \left(\frac{m_1}{2\pi kT} \right)^{3/2} \exp(-m_1 C_1^2/2kT), \\ f_2^{(0)} &= n_2 \left(\frac{m_2}{2\pi kT} \right)^{3/2} \exp(-m_2 C_2^2/2kT), \end{aligned} \right\} \quad \dots \quad (3.2)$$

where k is Boltzmann's constant and T the kinetic temperature of the gas. It is assumed here that the electric field is not so large as to warrant the retention of the term in \mathbf{E} on the left side of (3.1). To this approximation we have

$$\left. \begin{aligned} \mathcal{A}^{(0)}\bar{\phi}_1 &= \mathcal{A}^{(0)}\bar{\phi}_2 = 0, & \mathbf{j}_1^{(0)} &= 0, & \text{etc.,} \\ \mathbf{p}_1^{(0)} &= p_1 \mathbf{U}, & \mathbf{p}_2^{(0)} &= p_2 \mathbf{U}, \end{aligned} \right\} \quad \dots \quad (3.3)$$

where p_1, p_2 are the 'hydrostatic' partial pressures and \mathbf{U} the unit tensor.

To obtain the next approximations $f_1^{(0)} + f_1^{(1)}, f_2^{(0)} + f_2^{(1)}$ we proceed as indicated in C. & C., § 18.6, p. 345. For (2.1) write

$$\begin{aligned} \frac{\partial f_1^{(1)}}{\partial t} + \mathbf{c}_0 \cdot \frac{\partial f_1^{(0)}}{\partial \mathbf{r}} + \mathbf{C}_1 \cdot \frac{\partial f_1^{(0)}}{\partial \mathbf{r}} + \left\{ \left(\mathbf{F} + \frac{e_1}{m_1} (\mathbf{E} + \mathbf{c}_0 \wedge \mathbf{H}) - \frac{D\mathbf{c}_0}{Dt} \right) \right\} \cdot \frac{\partial f_1^{(0)}}{\partial \mathbf{C}_1} \\ + \frac{e_1}{m_1} \mathbf{C}_1 \wedge \mathbf{H} \cdot \frac{\partial f_1^{(1)}}{\partial \mathbf{C}_1} - \frac{\partial f_1^{(0)}}{\partial \mathbf{C}_1} \mathbf{C}_1 \cdot \frac{\partial}{\partial \mathbf{r}} \mathbf{c}_0 \\ = -J_1(f_1^{(0)}f_1^{(1)}) - J_1(f_1^{(1)}f_2^{(0)}) - J_{12}(f_1^{(0)}f_2^{(1)}) - J_{12}(f_1^{(1)}f_2^{(0)}). \end{aligned} \quad \dots \quad (3.4)$$

Then, if

$$f_1^{(1)} = f_1^{(0)}\Phi_1^{(1)}, \quad f_2^{(1)} = f_2^{(0)}\Phi_2^{(1)}, \quad \dots \quad (3.5)$$

the right side of (3.4) may be written (C. & C., p. 330)

$$-n_1^2 I_1(\Phi_1^{(1)}) - n_1 n_2 I_{12}(\Phi_1^{(1)} + \Phi_2^{(1)}). \quad \dots \quad (3.6)$$

The second approximation to the equation of change of $\bar{\phi}_1$ follows directly. Corresponding to (3.6), the rate of change due to collisions is given by

$$n_1 \mathcal{A}^{(1)} \bar{\phi}_1 = -n_1^2 [\phi, \Phi^{(1)}]_1 - n_1 n_2 [\phi_1, \Phi_1^{(1)} + \Phi_2^{(1)}]_{12}, \quad \dots \quad (3.7)$$

in terms of the integrals defined in C. & C., § 4.4. By adding a similar equation for $\bar{\phi}_2$ we obtain for the right side

$$n_1 \mathcal{A}^{(1)} \bar{\phi}_1 + n_2 \mathcal{A}^{(1)} \bar{\phi}_2 = -n_1 n_2 \{\phi, \Phi^{(1)}\}. \quad \dots \quad (3.8)$$

In particular, the second approximation to (2.5) is

$$\frac{\partial \mathbf{j}^{(1)}}{\partial t} - \frac{\rho_1 \rho_2}{\rho} \left(\frac{e_1}{m_1 \rho_1} + \frac{e_2}{m_2 \rho_2} \right) \mathbf{j}^{(1)} \wedge \mathbf{H} - \mathcal{A}^{(1)} \mathbf{j} = -p \left(\frac{e_1}{m_1} - \frac{e_2}{m_2} \right) \mathbf{d}_{12}, \quad (3.9)$$

where

$$\mathbf{d}_{12} = \frac{\partial(n_1/n)}{\partial \mathbf{r}} - \frac{\rho_1 \rho_2}{p \rho} \left\{ \left(\frac{e_1}{m_1} - \frac{e_2}{m_2} \right) (\mathbf{E} + \mathbf{c}_0 \wedge \mathbf{H}) - \left(\frac{1}{m_1} - \frac{1}{m_2} \right) \frac{1}{n} \frac{\partial p}{\partial \mathbf{r}} \right\}, \quad (3.10)$$

$$\mathcal{A}^{(1)} \mathbf{j} = -n_1 n_2 \{e \mathbf{C}, \Phi^{(1)}\}, \quad \dots \quad (3.11)$$

and

$$\mathbf{j}^{(1)} = \mathbf{j}_1^{(1)} + \mathbf{j}_2^{(1)}$$

$$= \int f_1^{(0)} \Phi_1^{(1)} e_1 \mathbf{C}_1 d\mathbf{c}_1 + \int f_2^{(0)} \Phi_2^{(1)} e_2 \mathbf{C}_2 d\mathbf{c}_2. \quad \dots \quad (3.12)$$

Here we have used the results

$$\left. \begin{aligned} p_1 &= (n_1/n)p, & p_2 &= (n_2/n)p, \\ \text{where } n_1 + n_2 &= n \text{ and } p = nkT. \end{aligned} \right\} \quad \dots \quad (3.13)$$

The corresponding approximations* to the equations of conservation, mass motion and thermal energy are

$$\frac{Dn_1}{Dt} + n_1 \frac{\partial}{\partial \mathbf{r}} \cdot \mathbf{c}_0 = 0, \quad \frac{Dn_2}{Dt} + n_2 \frac{\partial}{\partial \mathbf{r}} \cdot \mathbf{c}_0 = 0, \quad \dots \quad (3.14)$$

$$\rho \frac{D\mathbf{c}_0}{Dt} = \rho \mathbf{F} + \rho_e (\mathbf{E} + \mathbf{c}_0 \wedge \mathbf{H}) + \mathbf{j}^{(1)} \wedge \mathbf{H} - \frac{\partial p}{\partial \mathbf{r}}, \quad \dots \quad (3.15)$$

and

$$\frac{3}{2} kn \frac{DT}{Dt} = -p \frac{\partial}{\partial \mathbf{r}} \cdot \mathbf{c}_0, \quad \dots \quad (3.16)$$

where

$$\rho_e = n_1 e_1 + n_2 e_2. \quad \dots \quad (3.17)$$

§ 4. THE DETERMINATION OF $\mathcal{A}^{(1)}\mathbf{j}$

Since momentum is conserved in encounters between like molecules,

$$\mathbf{C}_1 + \mathbf{C} - \mathbf{C}'_1 - \mathbf{C}' = 0, \quad \dots \quad (4.1)$$

whence, from (2.3) and (3.7)

$$\left. \begin{aligned} \mathcal{A}^{(1)}\mathbf{j}_1 &= -n_1 n_2 [e_1 \mathbf{C}_1, \quad \Phi_1^{(1)} + \Phi_2^{(1)}]_{12}, \\ \mathcal{A}^{(1)}\mathbf{j}_2 &= -n_1 n_2 [e_2 \mathbf{C}_2, \quad \Phi_1^{(1)} + \Phi_2^{(1)}]_{12}, \end{aligned} \right\} \quad \dots \quad (4.2)$$

so that $\mathcal{A}^{(1)}\mathbf{j} = -n_1 n_2 [e_1 \mathbf{C}_1 + e_2 \mathbf{C}_2, \quad \Phi_1^{(1)} + \Phi_2^{(1)}]_{12}$.

It can be shown that the functions $\Phi_1^{(1)}, \Phi_2^{(1)}$ are of the form (C. & C., p. 331)

$$\left. \begin{aligned} \Phi_1^{(1)} &= -\mathbf{A}_1 \cdot \frac{\partial \ln T}{\partial \mathbf{r}} - \mathbf{B}_1 : \frac{\partial}{\partial \mathbf{r}} \mathbf{c}_0 - n \mathbf{D}_1 \cdot \mathbf{d}_{12}, \\ \Phi_2^{(1)} &= -\mathbf{A}_2 \cdot \frac{\partial \ln T}{\partial \mathbf{r}} - \mathbf{B}_2 : \frac{\partial}{\partial \mathbf{r}} \mathbf{c}_0 - n \mathbf{D}_2 \cdot \mathbf{d}_{12}, \end{aligned} \right\} \quad \dots \quad (4.3)$$

where the vectors \mathbf{A} and \mathbf{D} are linear and the tensors \mathbf{B} non-divergent dyadic functions of the vectors $\mathbf{C}, \mathbf{C} \wedge \mathbf{H}, (\mathbf{C} \wedge \mathbf{H}) \wedge \mathbf{H}$, whose coefficients are functions of the scalars C and H . They are subject to the conditions of solubility (C. & C., p. 142)

$$\left. \begin{aligned} \int f_1^{(0)} m_1 \mathbf{C}_1 \cdot \mathbf{A}_1 d\mathbf{c}_1 + \int f_2^{(0)} m_2 \mathbf{C}_2 \cdot \mathbf{A}_2 d\mathbf{c}_2 &= 0, \\ \int f_1^{(0)} m_1 \mathbf{C}_1 \cdot \mathbf{D}_1 d\mathbf{c}_1 + \int f_2^{(0)} m_2 \mathbf{C}_2 \cdot \mathbf{D}_2 d\mathbf{c}_2 &= 0. \end{aligned} \right\} \quad \dots \quad (4.4)$$

On substitution from (4.3) into (4.2), the terms in \mathbf{B}_1 and \mathbf{B}_2 give zero contributions (see Appendix I) so that

$$\mathcal{A}^{(1)}\mathbf{j}_1 = n_1 n_2 \left[e_1 \mathbf{C}_1, (\mathbf{A}_1 + \mathbf{A}_2) \cdot \frac{\partial \ln T}{\partial \mathbf{r}} + n(\mathbf{D}_1 + \mathbf{D}_2) \cdot \mathbf{d}_{12} \right]_{12}, \text{ etc.} \quad (4.5)$$

Also in (3.12) the terms in \mathbf{B}_1 and \mathbf{B}_2 are of odd degree in the components of \mathbf{C}_1 and \mathbf{C}_2 . Hence

$$\mathbf{j}_1^{(1)} = - \int f_1^{(0)} \left(\mathbf{A}_1 \cdot \frac{\partial \ln T}{\partial \mathbf{r}} + n \mathbf{D}_1 \cdot \mathbf{d}_{12} \right) e_1 \mathbf{C}_1 d\mathbf{c}_1, \text{ etc.} \quad \dots \quad (4.6)$$

* These equations are not identical with those for static fields in C. & C., pp. 330, 331; there the symbol D_0/Dt in effect denotes $\mathbf{c}_0 \cdot (\partial/\partial \mathbf{r})$.

It follows that, to the present order of approximation, viscous forces have no effect on the conduction current. In C. & C., ch. 18 they are simply neglected.

The novel feature of the present investigation is that the evaluation of the vector functions \mathbf{A} and \mathbf{D} is avoided by seeking a direct relation between $\mathcal{A}^{(1)}\mathbf{j}$ and $\mathbf{j}^{(1)}$, as suggested by the 'free-path' result (2.8). We may write

$$\left. \begin{aligned} \mathbf{A}_1 &= \mathbf{a}_1 \cdot \mathbf{C}_1, & \mathbf{A}_2 &= \mathbf{a}_2 \cdot \mathbf{C}_2, \\ \mathbf{D}_1 &= \mathbf{d}_1 \cdot \mathbf{C}_1, & \mathbf{D}_2 &= \mathbf{d}_2 \cdot \mathbf{C}_2, \end{aligned} \right\} \quad \dots \quad (4.7)$$

where the tensors \mathbf{a} and \mathbf{d} involve the components of \mathbf{H} and scalar functions of C and H . The latter functions are conveniently represented by expansions in terms of Sonine polynomials of argument C^2 . We restrict our consideration to the first sub-approximation which retains only the first terms of these expansions, so that the \mathbf{a} and \mathbf{d} become independent of \mathbf{C} . Cowling (1945) states that this procedure is adequate, except when ω_H/ν (see below) is small. Then (4.5) gives (see Appendix II)

$$\begin{aligned} \mathcal{A}^{(1)}\mathbf{j}_1 &= n_1 n_2 \left[e_1 \mathbf{C}_1, (\mathbf{a}_1 \cdot \mathbf{C}_1 + \mathbf{a}_2 \cdot \mathbf{C}_2) \cdot \frac{\partial \ln T}{\partial \mathbf{r}} \right. \\ &\quad \left. + n(\mathbf{d}_1 \cdot \mathbf{C}_1 + \mathbf{d}_2 \cdot \mathbf{C}_2) \cdot \mathbf{d}_{12} \right]_2 \\ &= \frac{1}{3} n_1 n_2 e_1 \left(\frac{\partial \ln T}{\partial \mathbf{r}} \cdot \mathbf{a}_1 + n \mathbf{d}_{12} \cdot \mathbf{d}_1 \right) [\mathbf{C}_1, \mathbf{C}_1]_{12} \\ &\quad + \frac{1}{3} n_1 n_2 e_1 \left(\frac{\partial \ln T}{\partial \mathbf{r}} \cdot \mathbf{a}_2 + n \mathbf{d}_{12} \cdot \mathbf{d}_2 \right) [\mathbf{C}_1, \mathbf{C}_2]_{12}, \quad \dots \quad (4.8) \end{aligned}$$

where the products in the integrals represented by the brackets are scalar. Since in an encounter between unlike molecules

$$m_1(\mathbf{C}_1 - \mathbf{C}'_1) + m_2(\mathbf{C}_2 - \mathbf{C}'_2) = 0, \quad \dots \quad (4.9)$$

$$m_1[\mathbf{C}_1, \mathbf{C}_1]_{12} + m_2[\mathbf{C}_1, \mathbf{C}_2]_{12} = 0. \quad \dots \quad (4.10)$$

Further, since only terms of even degree in the components of \mathbf{C} survive integration, (4.4) gives

$$\frac{1}{3} \mathbf{U} : \mathbf{a}_1 \int f_1^{(0)} m_1 C_1^2 d\mathbf{c}_1 + \frac{1}{3} \mathbf{U} : \mathbf{a}_2 \int f_2^{(0)} m_2 C_2^2 d\mathbf{c}_2 = 0,$$

$$\frac{1}{3} \mathbf{U} : \mathbf{d}_1 \int f_1^{(0)} m_1 C_1^2 d\mathbf{c}_1 + \frac{1}{3} \mathbf{U} : \mathbf{d}_2 \int f_2^{(0)} m_2 C_2^2 d\mathbf{c}_2 = 0,$$

whence

$$(n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2) : \mathbf{U} = 0,$$

$$(n_1 \mathbf{d}_1 + n_2 \mathbf{d}_2) : \mathbf{U} = 0.$$

Thus the conditions (4.4) are satisfied by taking

$$\left. \begin{aligned} n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2 &= 0, \\ n_1 \mathbf{d}_1 + n_2 \mathbf{d}_2 &= 0. \end{aligned} \right\} \quad \dots \quad (4.11)$$

Then substitution from (4.10) and (4.11) into (4.8) gives

$$\mathcal{A}^{(1)}\mathbf{j}_1 = -\frac{1}{3} \rho \frac{n_1 e_1}{m_1} \left(\frac{\partial \ln T}{\partial \mathbf{r}} \cdot \mathbf{a}_1 + n \mathbf{d}_{12} \cdot \mathbf{d}_1 \right) [\mathbf{C}_1, \mathbf{C}_2]_{12}. \quad (4.12)$$

Similarly, (4.6) gives

$$\begin{aligned}\mathbf{j}_1^{(1)} &= -\frac{1}{3} \left(\frac{\partial \ln T}{\partial \mathbf{r}} \cdot \mathbf{a}_1 + n \mathbf{d}_{12} \cdot \mathbf{d}_1 \right) \int f_1^{(0)} e_1 C_1^2 d\mathbf{c}_1 \\ &= -kT \frac{n_1 e_1}{m_1} \left(\frac{\partial \ln T}{\partial \mathbf{r}} \cdot \mathbf{a}_1 + n \mathbf{d}_{12} \cdot \mathbf{d}_1 \right). \quad \dots \dots \dots \quad (4.13)\end{aligned}$$

Thus we have

$$\Delta^{(1)} \mathbf{j}_1 = -\nu \mathbf{j}_1^{(1)}, \quad \dots \dots \dots \quad (4.14)$$

where

$$\nu = -\frac{\rho}{3kT} [\mathbf{C}_1, \mathbf{C}_2]_{12}. \quad \dots \dots \dots \quad (4.15)$$

The same relation holds between $\Delta^{(1)} \mathbf{j}_2$ and $\mathbf{j}_2^{(1)}$.

C. & C., §§ 9.8, 9.81 give

$$[\mathbf{C}_1, \mathbf{C}_2]_{12} = -\frac{3(kT)^2}{nm_1 m_2 D_{12}}, \quad \dots \dots \dots \quad (4.16)$$

where D_{12} is the coefficient of mutual diffusion, given to the same order of approximation as in (4.12) and (4.13). Thus

$$\nu = \frac{\rho kT}{nm_1 m_2 D_{12}}. \quad \dots \dots \dots \quad (4.17)$$

It is identical with the parameter $1/\tau$ on C. & C., p. 334, which is a weighted mean of the ion and electron collision frequencies. With $m_2 \ll m_1$, $\nu = (4/3)\nu_2$, where ν_2 is the electron collision frequency. This modification of the 'free-path' result (2.8) confirms Smerd and Westfold's (1949) explanation of a discrepancy of 4/3 between the formula for the absorption coefficient as obtained from the usual Lorentz theory and that derived from the formula for the emissivity obtained by velocity-distribution methods.

C. & C., p. 179 gives

$$D_{12} = \frac{3}{16n} \left(\frac{2kT}{\pi m_0 M_1 M_2} \right)^{1/2} \left(\frac{2kT}{e_1 e_2} \right)^2 / \ln(1 + v_{01}^2), \quad \dots \quad (4.18)$$

where

$$m_0 = m_1 + m_2, \quad M_1 = m_1/m_0, \quad M_2 = m_2/m_0, \quad \dots \quad (4.19)$$

e_1, e_2 are in electrostatic units and $\pi/2 \operatorname{artan} v_{01}$ is the smallest deflection in a Coulomb interaction that is reckoned as an encounter. Thus we arrive at the formula

$$\nu = \frac{4\pi\rho e_1^2 e_2^2 \ln(1 + v_{01}^2)}{3m_c kT (2\pi m_0 M_1 M_2 kT)^{1/2}}, \quad \dots \dots \dots \quad (4.20)$$

which should be compared with (3.11) of Smerd and Westfold (1949).

The value to be assigned to v_{01} has been discussed in detail elsewhere (Westfold 1951 b, § 4.5). The relative orbit of the interacting molecules is determined by the relative speed g before the encounter and the collision parameter b . Encounters for which $b > d$, the mean

distance* between pairs of molecules, are disregarded, as they can no longer be regarded as binary. Further, those for which $b > g/\omega$ are reckoned as being irrelevant to radiation of frequency $f = \omega/2\pi$. Thus, averaging over all velocities, we are led to take

$$\left. \begin{aligned} v_{01} &= -\frac{4kTd}{e_1 e_2} \quad \text{for } \frac{1}{2}\pi f d \left(\frac{2\pi m_0 M_1 M_2}{kT} \right)^{1/2} < 1, \\ v_{01} &= -\frac{15(kT)^{3/2}}{4e_1 e_2 f (2\pi m_0 M_1 M_2)^{1/2}} \quad \text{for } \frac{1}{2}\pi f d \left(\frac{2\pi m_0 M_1 M_2}{kT} \right)^{1/2} > 1. \end{aligned} \right\} \quad (4.21)$$

The former value was obtained by Smerd and Westfold (1949). It is applicable to the solar atmosphere and the H II clouds of interstellar space. The latter value is applicable to the diffuse H II regions.

They are equivalent to those adopted by Denisse (1950), but exceed those of Burkhardt, Elwert and Unsöld (1948), who restrict their consideration to small deflections, by the factors 5.8, 7.0 respectively.

§ 5. SOME SPECIAL CASES

On introducing the magnetic field parameter

$$\omega_H = \frac{\rho_1 \rho_2}{\rho} \left(\frac{e_1}{m_1 \rho_1} + \frac{e_2}{m_2 \rho_2} \right) \mathbf{H}, \quad \dots \quad (5.1)$$

whose magnitude is a weighted mean of the ion and electron angular gyro-frequencies, and substituting from (4.14), our approximation (3.9) to the generalized Lorentz equation becomes

$$\frac{\partial \mathbf{j}}{\partial t} + \omega_H \mathbf{j} + \mathbf{j} \mathbf{j} = -p \left(\frac{e_1}{m_1} - \frac{e_2}{m_2} \right) \mathbf{d}_{12}. \quad \dots \quad (5.2)$$

The equations of the electromagnetic field with (5.2), (3.14)–(3.16) provide a sufficient and reliable set for investigating the interactions between an ionized gas in motion in the presence of imposed gravitational, electric and magnetic fields, and a radiation field.

They invite comparison with the equations of Bailey's (1948) electro-magneto-ionic theory, which are expressed in terms of the mean velocities $\bar{\mathbf{c}}_1, \bar{\mathbf{c}}_2$. Bailey's fundamental equations are in accord with our set except where he takes collisional effects as being proportional to $\bar{\mathbf{c}}_1, \bar{\mathbf{c}}_2$ instead of to the mean peculiar velocities $\bar{\mathbf{C}}_1, \bar{\mathbf{C}}_2$. However, in applications collisional effects are largely neglected.

If \mathbf{c}_0 is small compared with the velocity of sound, it is known that the gas behaves as if it were incompressible. Then from (3.14) we have

$$\frac{\partial}{\partial \mathbf{r}} \cdot \mathbf{c}_0 = 0, \quad \dots \quad (5.3)$$

whence, from (3.16),

$$\frac{DT}{Dt} = 0. \quad \dots \quad (5.4)$$

* It might be argued that the Debye distance $\lambda_D = \{kT/4\pi (n_1 e_1^2 + n_2 e_2^2)\}^{1/2}$, beyond which the shielding effect of the surrounding molecules is negligible, is a more appropriate value for the limiting distance. However, in the solar atmosphere and in the H II regions we find $\lambda_D > d$, so that d is to be preferred.

Thus, to a first approximation, the temperature of a volume element of the gas, as well as its density, remains constant during its motion.

In the magneto-ionic and Lorentz theories the gas is at rest, save for possible oscillatory motions induced by the electromagnetic field. In the expression (3.10) for \mathbf{d}_{12} , gradients of density and pressure are neglected, \mathbf{E} represents the electric vector of the electromagnetic field and \mathbf{H} the imposed magnetic field. Substitution into (5.2) gives

$$\frac{\partial \mathbf{j}}{\partial t} + \omega_H \wedge \mathbf{j} + \nu \mathbf{j} = \epsilon_v \omega_0^2 (\mathbf{E} + c_0 \wedge \mathbf{H}), \quad \dots \dots \dots \quad (5.5)$$

where

$$\omega_0^2 = \frac{\rho_1 \rho_2}{\rho \epsilon_v} \left(\frac{e_1}{m_1} - \frac{e_2}{m_2} \right)^2 \quad \dots \dots \dots \quad (5.6)$$

is a quadratic function of the ion and electron angular plasma frequencies. In rationalized e.m.u.* ϵ_v is the permittivity of free space. Neglecting also the pressure gradient and non-linear terms in (3.15) we get

$$\rho \frac{\partial \mathbf{c}_0}{\partial t} = \mathbf{j} \wedge \mathbf{H}. \quad \dots \dots \dots \quad (5.7)$$

The elimination of \mathbf{c}_0 between (5.5) and (5.7) results in an anisotropic conductivity relation between \mathbf{j} and \mathbf{E} (cf. Westfold 1949), which with Maxwell's equations leads to formulae for the refractive index and absorption coefficient for radiation of a given frequency $\omega/2\pi$. The effect of thus allowing for the induced motion of the gas is to introduce additional small terms of order $\rho_2 \omega_H^2 / \rho_1 \omega^2$ into the usual magneto-ionic formulae. When $\mathbf{H}=0$ (5.5) reduces to the isotropic Lorentz form.

When only static fields are considered we put $\partial \mathbf{j} / \partial t = 0$ in (5.2) and it is then not difficult to recover the transverse conductivity relation† of C. & C., p. 335, for which $\omega_H \cdot \mathbf{d}_{12} = 0$.

§ 6. THE CASE OF A SLIGHTLY IONIZED GAS

We have already noted that a binary ionized gas is subject to the Lorentzian condition, $m_2 \ll m_1$, where m_1 and m_2 are the masses of an ion and an electron respectively. It then follows that the ionic component of the current is negligible compared with the electron component so that

$$\mathbf{j} \sim \mathbf{j}_2. \quad \dots \dots \dots \quad (6.1)$$

Also,

$$\omega_H \sim \omega_{H2} = e_2 \mathbf{H} / m_2 \quad \dots \dots \dots \quad (6.2)$$

the electron angular gyro-frequency, and

$$\omega_0^2 \sim \omega_{02}^2 = n_2 e_2^2 / m_2 \epsilon_v, \quad \dots \dots \dots \quad (6.3)$$

* With these units it is customary to replace \mathbf{H} by the magnetic induction \mathbf{B} in such equations as occur in this paper.

† This result is expressed in terms of the velocity of diffusion $\overline{\mathbf{C}_1} - \overline{\mathbf{C}_2}$, which is related to \mathbf{j} by the formula

$$\mathbf{j} = (\rho_1 \rho_2 / \rho) (e_1 / m_1 - e_2 / m_2) (\overline{\mathbf{C}_1} - \overline{\mathbf{C}_2}).$$

the square of the electron angular plasma frequency. Thus (5.2) and (3.10) are effectively given by

$$\frac{\partial \mathbf{j}}{\partial t} + \omega_{H2} \mathbf{j} + \nu \mathbf{j} = \epsilon_v \omega_{02}^2 (\mathbf{E} + \mathbf{c}_0 \wedge \mathbf{H}) - \frac{e_2}{m_2} \frac{\partial p_2}{\partial \mathbf{r}}, \quad \dots \quad (6.4)$$

which may be compared with (2.5).

It follows further that (6.4) may be applied to a gas which is only slightly ionized, provided that the damping factor ν refers now to collisions between electrons and the neutral molecules which comprise the bulk of the gas. In this case the diffusion coefficient is given by (C. & C., §§ 9.81, 10.22)

$$D_{12} = \frac{3}{16n\sigma_{12}^2} \left(\frac{2kT}{\pi m_0 M_1 M_2} \right)^{1/2}, \quad \dots \quad (6.5)$$

where the subscript 1 now refers to the neutral molecules and σ_{12} is the radius of the cross-section for collisions between electrons and neutral molecules. Substituting into (4.17) and applying the Lorentzian condition, we get

$$\nu = \frac{8n_1\sigma_{12}^2}{3} \left(\frac{2\pi kT}{m_2} \right)^{1/2} \quad \dots \quad (6.6)$$

This result is applicable to the lower layers of the ionosphere and the H I clouds of interstellar space. Again, it differs from the 'free-path' value by the factor 4/3.

ACKNOWLEDGMENT

The author has had the benefit of discussions on this subject with Professor S. Chapman and Professor T. G. Cowling.

APPENDIX I

To show that

$$[\mathbf{C}_1, \mathbf{B}_r : \mathbf{a}]_{12} = 0, \quad r = 1, 2,$$

where \mathbf{a} is a constant tensor.

From C. & C., § 4.4

$$n_1 n_2 [\mathbf{C}_1, \mathbf{B}_r : \mathbf{a}]_{12} = \iiint f_1^{(0)} f_2^{(0)} (\mathbf{C}_1 - \mathbf{C}'_1) \mathbf{B}_r : \mathbf{a} k_{12} d\mathbf{k} d\mathbf{c}_1 d\mathbf{c}_2.$$

We make the transformation

$$\mathbf{C}_1 = \mathbf{G}_0 - M_2 \mathbf{g}, \quad \mathbf{C}_2 = \mathbf{G}_0 + M_1 \mathbf{g},$$

where \mathbf{G}_0 is the velocity, relative to the mass velocity, of the centre of mass of a pair of interacting molecules, and \mathbf{g} the velocity of the molecule 2 relative to the molecule 1. Then

$$\begin{aligned} d\mathbf{c}_1 d\mathbf{c}_2 &= d\mathbf{C}_1 d\mathbf{C}_2 = d\mathbf{G}_0 d\mathbf{g}, \\ m_1 C_1^2 + m_2 C_2^2 &= m_0 (G_0^2 + M_1 M_2 g^2), \end{aligned}$$

and

$$n_1 n_2 [\mathbf{C}_1, \mathbf{B}_r : \mathbf{a}]_{12} = -M_2 \iiint f_1^{(0)} f_2^{(0)} (\mathbf{g} - \mathbf{g}') \mathbf{B}_r : \mathbf{a} k_{12} d\mathbf{k} d\mathbf{G}_0 d\mathbf{g}.$$

All the elements of the tensors \mathbf{B}_r are of even degree in the components of \mathbf{G}_0 and \mathbf{g} . Since the components of \mathbf{g}' , the relative velocity after collision, are linear functions of the components of \mathbf{g} (\mathbf{g}' represents a rotation of \mathbf{g}), $\mathbf{g} - \mathbf{g}'$ is of odd and the rest of the integrand of even degree in the components of \mathbf{G}_0 and \mathbf{g} . Hence all three components of the vector represented by the integral $[\mathbf{C}_1, \mathbf{B}_r : \mathbf{a}]_{12}$ vanish.

APPENDIX II

To show that

$$[\mathbf{C}_1, \mathbf{a} : \mathbf{C}_r \alpha]_{12} = \frac{1}{3} \alpha \cdot \mathbf{a} [\mathbf{C}_1, \mathbf{C}_r]_{12}, \quad r=1, 2,$$

where \mathbf{a} is a constant tensor and α a constant vector.

As in Appendix I,

$$\begin{aligned} n_1 n_2 [\mathbf{C}_1, \mathbf{a} : \mathbf{C}_r \alpha]_{12} &= \iiint f_1^{(0)} f_2^{(0)} (\mathbf{C}_1 - \mathbf{C}'_1) \mathbf{a} : \mathbf{C}_r \alpha k_{12} d\mathbf{k} d\mathbf{c}_1 d\mathbf{c}_2, \\ &= \iiint f_1^{(0)} f_2^{(0)} \mathbf{C}_1 \mathbf{a} : (\mathbf{C}_r - \mathbf{C}'_r) \alpha k_{12} d\mathbf{k} d\mathbf{c}_1 d\mathbf{c}_2 \end{aligned}$$

by virtue of the commutative property of the brackets. For $r=1$, the former integral transforms into

$$\begin{aligned} &- M_2 \iiint f_1^{(0)} f_2^{(0)} (\mathbf{g} - \mathbf{g}') \mathbf{a} : (\mathbf{G}_0 - M_2 \mathbf{g}) \alpha k_{12} d\mathbf{k} d\mathbf{G}_0 d\mathbf{g}, \\ &= M_2^2 \iiint f_1^{(0)} f_2^{(0)} (\mathbf{g} - \mathbf{g}') \mathbf{a} : \mathbf{g} \alpha k_{12} d\mathbf{k} d\mathbf{G}_0 d\mathbf{g}, \end{aligned}$$

since the integral is of odd degree in the components of \mathbf{G}_0 . Similarly the latter integral becomes

$$M_2^2 \iiint f_1^{(0)} f_2^{(0)} \mathbf{g} \mathbf{a} : (\mathbf{g} - \mathbf{g}') \alpha k_{12} d\mathbf{k} d\mathbf{G}_0 d\mathbf{g}.$$

Thus

$$\iiint f_1^{(0)} f_2^{(0)} \alpha \cdot \mathbf{a} \cdot (\mathbf{g} \mathbf{g}' - \mathbf{g}' \mathbf{g}) k_{12} d\mathbf{k} d\mathbf{G}_0 d\mathbf{g} = 0.$$

But since \mathbf{g}' is not parallel to \mathbf{g} , $\mathbf{g} \mathbf{g}' \neq \mathbf{g}' \mathbf{g}$; hence only the diagonal terms of the dyadic $\mathbf{g} \mathbf{g}'$ can make non-zero contributions to the original integral. The same is true of the dyadic $\mathbf{g} \mathbf{g}$ since the non-diagonal terms are of odd degree in the components of \mathbf{g} . Thus

$$\begin{aligned} n_1 n_2 [\mathbf{C}_1, \mathbf{a} : \mathbf{C}_1 \alpha]_{12} &= \frac{1}{3} M_2^2 \alpha \cdot \mathbf{a} \iiint f_1^{(0)} f_2^{(0)} (\mathbf{g} - \mathbf{g}') \cdot \mathbf{g} k_{12} d\mathbf{k} d\mathbf{G}_0 d\mathbf{g} \\ &= \frac{1}{3} \alpha \cdot \mathbf{a} n_1 n_2 [\mathbf{C}_1, \mathbf{C}_1]_{12}. \end{aligned}$$

It may similarly be proved that

$$[\mathbf{C}_1, \mathbf{a} : \mathbf{C}_2 \alpha]_{12} = \frac{1}{3} \alpha \cdot \mathbf{a} [\mathbf{C}_1, \mathbf{C}_2]_{12}.$$

REFERENCES

BAILEY, V. A., 1948, *Aust. J. Sci. Res. A*, **1**, 351-9.
 BHATNAGAR, P. L., KROOK, M., and MENZEL, D. H., 1952, *Preliminary Report of the Committee on Dynamics of Ionized Media*. U.R.S.I. Report.
 BOHM, D., and GROSS, E. P., 1949, *Phys. Rev.*, **75**, 1851-64.
 BURKHARDT, G., ELWERT, G., and UNSÖLD, A., 1948, *Z. Astrophys.*, **25**, 310-4.
 BURKHARDT, G., and SCHLÜTER, A., 1949, *Z. f. Astrophys.*, **26**, 295-304.

CHAPMAN, S., and COWLING, T. G., 1939, *The Mathematical Theory of Non-uniform Gases* (Cambridge : University Press).

COHEN, R. S., SPITZER, L., and ROUTLY, P. McR., 1950, *Phys. Rev.*, **80**, 230-8.

COWLING, T. G., 1945, *Proc. Roy. Soc. A*, **183**, 453-79.

DENISSE, J. F., 1950, *J. de Phys. et le Radium*, **11**, 164-71.

JAEGER, J. C., and WESTFOLD, K. C., 1949, *Aust. J. Sci. Res. A*, **2**, 322-34; 1950, *Ibid.*, **A**, **3**, 376-86.

MARTYN, D. F., 1947, *Nature, Lond.*, **159**, 26-7.

SHKLOVSKY, I. S., 1946, *Astronom. J. U. S. S. R.*, **23**, 333-47.

SMERD, S. F., and WESTFOLD, K. C., 1949, *Phil. Mag.*, **40**, 831-48.

WESTFOLD, K. C., 1949, *Aust. J. Sci. Res. A*, **2**, 169-83; 1951 a, *Proceedings of a Conference on the Dynamics of Ionized Media* (London); 1951 b, *D. Phil. Thesis*, Oxford.

LXXVI. *Conditions for the Occurrence of Electrical Discharges
in Astrophysical Systems*

By J. W. DUNGEY

School of Physics, The University of Sydney, Australia*

[Received November 14, 1952, revised March 11, 1953]

SUMMARY

Discharges are shown to be a possible source of high energy particles, if the current density is very large. The growth of the current density is discussed using the fact that the magnetic lines of force are approximately frozen into the ionized gas. It is shown that discharges are unlikely to occur anywhere except at neutral points of the magnetic field. Neutral points are found to be unstable in such a way that a small perturbation will start a discharge in a time of the order of the characteristic time of the system. Such discharges may account for aurorae, and may also occur in solar flares and the interstellar gas.

§ 1. INTRODUCTION

THE possibility of the occurrence of electric discharges in astrophysical systems is important as an obvious source of high energy particles, if the accelerating voltage is large enough. The existence of large 'potential differences' in rotating magnetic stars has been pointed out by Alfvén and others (Alfvén 1950), but this is not a sufficient condition for a discharge to occur, as will be seen later. Further, the possibility of the particles in a small region reaching very high energies by absorbing energy from a large surrounding region is of more interest than the moderate heating of the material in a large region.

In the following no particular system is discussed, but any system of interest can be described for our purposes as a large mass of ionized gas in a more or less complicated state of motion. A 'discharge' will be a region in which the electrons are accelerated to high energies by the electric field, so that all the electrons are moving in the same direction with large velocities. If we suppose that the electrons acquire relativistic energies, the current density is then of order nec , where n is the electron density. Now Maxwell's equations show that $c \operatorname{curl} \mathbf{H}/4\pi$ must be approximately equal to the current density and nec is found to be large when compared with the values of $c |\operatorname{curl} \mathbf{H}|/4\pi$ usually expected in astrophysical systems. For instance in the chromosphere n is about 10^{11} particles/cm³ which requires $|\operatorname{curl} \mathbf{H}| \sim 500$ gauss/cm and this is much

* Communicated by the Author.

larger than the values expected in sunspots. In the interstellar gas n is about one particle/cm³, and the interstellar magnetic field is generally supposed not to exceed 10⁻⁵ gauss; then in a discharge the field would have to change considerably in a distance of 20 metres. Consequently a discharge must be extremely thin in one direction. In order to examine how such a discharge can occur a general study of cosmic electrodynamics is required.

In this paper the conditions which can lead to the onset of a discharge will alone be discussed. The arguments to be used apply only when the current density is small compared with nec ; they are sufficient to determine the conditions under which the current density will grow, although the behaviour of the discharge, once started, is more complicated. For a discussion of the behaviour of discharges the reader is referred to Alfvén (1950).

§ 2. COSMIC ELECTRODYNAMICS

The fundamental variables of cosmic electrodynamics are the electric and magnetic fields, \mathbf{E} and \mathbf{H} , the charge and current densities, ρ and \mathbf{j} , the mass density μ , velocity \mathbf{u} and pressure p of the gas. Gaussian units are used. The fundamental equations are Maxwell's equations, the hydrodynamical equation

$$\mu d\mathbf{u}/dt = -\nabla p + \rho \mathbf{E} + \mathbf{j} \wedge \mathbf{H}/c, \dots \dots \dots \quad (1)$$

in which

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla$$

and Ohm's law, which can be written (Sweet 1949)

$$\mathbf{E} + (\mathbf{u} \wedge \mathbf{H})/c = \mathbf{j}/\sigma + \mathbf{j} \wedge \mathbf{H}/ne, \dots \dots \dots \quad (2)$$

where σ is the conductivity and the last term is the Hall electric field; Cowling (1945) obtains $\sigma = c^2 T^{3/2} / k$, where T is the temperature of the gas and k may be treated as a constant in these applications, equal to 6.8×10^{13} , so that

$$\sigma = 1.3 \times 10^7 T^{3/2} \text{ sec}^{-1}.$$

The natural approach to the problem by considering the effect of an 'applied' electric field is not convenient here owing to certain sources of confusion which will now be mentioned. Equation (2) is obtained by calculating $\partial \mathbf{j} / \partial t^*$ and should contain a term $(m/ne^2) \partial \mathbf{j} / \partial t$ on the right-hand side, where m is the electron mass; using Maxwell's equations this can be written as

$$-\frac{m}{4\pi ne^2} \left(c^2 \operatorname{curl} \operatorname{curl} \mathbf{E} + \frac{\partial^2 \mathbf{E}}{\partial t^2} \right)$$

$(mc^2/4\pi ne^2)^{1/2}$ is the 'electron plasma wavelength' and is usually small compared with the dimensions of astrophysical systems, so that the

* Equations (1) and (2) are obtained from the separate equations of motion for each kind of ion present.

omission of this term from (2) is justified. It should be included, however, if the effect of an applied electric field is considered. It is then seen that plasma oscillations must occur. Nor should it be supposed that these oscillations are damped out and that the current density changes in such a way as to satisfy (2) with the initial value of \mathbf{E} . The situation is controlled by induction effects. Suppose that the magnetic field vanishes initially and the applied electric field is uniform: $\text{curl } \mathbf{E}$ vanishes, hence \mathbf{H} continues to vanish, and $\partial \mathbf{E} / \partial t = -4\pi \mathbf{j} = -4\pi\sigma \mathbf{E}$. Consequently the electric field decays with a time constant $(4\pi\sigma)^{-1}$, or $6 \times 10^{-9} T^{-3/2}$ sec.

The growth of the magnetic field and current density depends on $\text{curl } \mathbf{E}$ not vanishing as in the theory of the skin effect, and as this sort of behaviour is controlled by induction it is better not to consider an applied electric field, but to study the behaviour of the gas, when eqn. (2) is satisfied all the time. The values of the variables at any time must then have arisen from past developments of the system. The high conductivity makes it possible to find a more successful approach, and this has been developed by several authors. (Walén 1947, Elsasser 1947, Dungey 1950). The expression for \mathbf{E} given by eqn. (2) determines $\partial \mathbf{H} / \partial t$: the contribution of each term can be estimated and it is found that for a system of astrophysical dimensions a good approximation is obtained if both terms on the right-hand side are neglected. In this approximation the rate of change of the magnetic field is given by

$$\partial \mathbf{H} / \partial t = -(\mathbf{u} \cdot \nabla) \mathbf{H} + (\mathbf{H} \cdot \nabla) \mathbf{u} - \mathbf{H} (\nabla \cdot \mathbf{u}) \dots \dots \quad (3)$$

and it can be shown that the magnetic flux, linked by a closed curve which moves with the gas, is a constant of the motion. It is then easy to see that the magnetic field is 'frozen into' the gas. This approximation is used in the following discussion of the possible initiation of a discharge.

It may be recalled here that when the magnitude of the current density was compared with $c \text{curl } \mathbf{H} / 4\pi$ in § 1, the displacement current was ignored. It can now be shown to be negligible. If the dimensions of the system are regarded as characterized by a length a and a velocity v , its characteristic time will be a/v and we may then put for instance $|\text{curl } \mathbf{H}| \sim |\mathbf{H}|/a$, $|\partial \mathbf{E} / \partial t| \sim v |\mathbf{E}|/a$. Then since $\mathbf{E} \approx \mathbf{H} \wedge \mathbf{u}/c$, we have $|\partial \mathbf{E} / \partial t| \sim v^2 |\mathbf{H}|/ac$ which can be neglected in comparison with $c |\text{curl } \mathbf{H}|$. Similarly in (1) $|\rho \mathbf{E}|$ is of order $v^2 |\mathbf{H}|^2/c^2 a$ and can therefore be neglected in comparison with $\mathbf{j} \wedge \mathbf{H}/c$. The electromagnetic force density can then be written $(\text{curl } \mathbf{H}) \wedge \mathbf{H}/4\pi$ or $(\mathbf{H} \cdot \nabla) \mathbf{H} - \frac{1}{2} \nabla |\mathbf{H}|^2/4\pi$, and can be represented as a tension $|\mathbf{H}|^2/8\pi$ per unit area along the lines of force and a lateral pressure of the same amount perpendicular to them.

§ 3. DISCHARGES IN A MAGNETIC FIELD

Since $|\text{curl } \mathbf{H}|$ needs to be very large when there is a discharge, and since the lines of force of the magnetic field are frozen in, it is essential to start with some magnetic field, and investigate how $|\text{curl } \mathbf{H}|$ can grow.

Consider first the possibility of a discharge occurring anywhere other than at a neutral point of the magnetic field. If the current density were to have a large component perpendicular to the magnetic field the electromagnetic force density would be large. Such a discharge might be expected to occur, when a shock wave travels across a magnetic field, but shock waves will not be discussed in this paper.

If the current density is parallel to the magnetic field, the lines of force are twisted like the strands of a cable. Because of the effective tension of the electromagnetic force we expect them to resist this twisting. A simple illustration is obtained in cylindrical coordinates ρ, ϕ, z by taking $H_\rho = 0$, $H_\phi = 8\pi J/c$, $H_z = H$; where H and J are constants; then $j_\rho = j_\phi = 0$, $j_z = J$. This represents a possible situation in the neighbourhood of a discharge, but it is necessary to consider what happens further away. The current lines must be closed and, unless they flow right round lines of force, they must cross the lines of force. Then there is a torque on the gas arising from the term $(H \cdot \nabla)H/4\pi$ in the force density, and directed so as to untwist the lines of force. Consequently the growth of the current density is opposed by the electromagnetic forces. This argument does not apply to the case when the current flows right round the lines of force, but then the lines of force are linked with each other a large number of times, if the current density is large. Now when lines of force are frozen into the gas, they cannot become linked during the course of the motion, and hence a large current density of this type cannot be a result of the motion. Consequently we do not expect discharges to occur except at neutral points of the magnetic field, which will now be discussed.

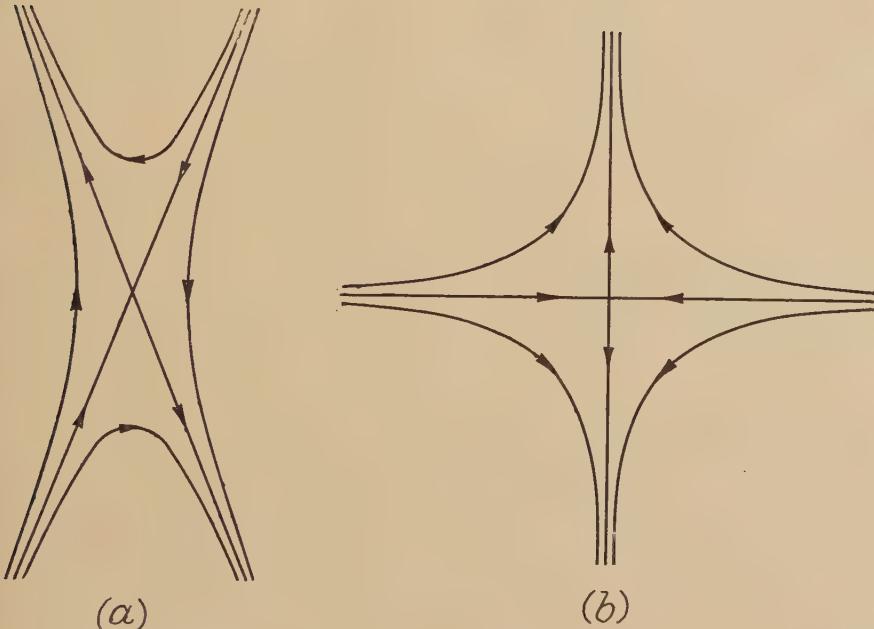
§ 4. NEUTRAL POINTS

Giovanelli (1947, 1948) and Hoyle (1949) have suggested that the neighbourhood of a neutral point is the seat of a discharge. Giovanelli points out that solar flares frequently occur in positions where a neutral point of the sunspot field is expected, and Hoyle has suggested an explanation of the origin of the aurora involving the same idea. We distinguish between two types of neutral point: X-type as in fig. 1 (a), and O-type which occurs at the centre of O-shaped lines of force. The possibility of discharges occurring elsewhere was rejected because the electromagnetic forces oppose the growth of the current density, but at an X-type neutral point the opposite situation occurs.

The magnetic field in the neighbourhood of a neutral point is described by the tensor $\partial H_i / \partial x_j$. The antisymmetrical part relates to $\text{curl } \mathbf{H}$ and therefore to \mathbf{j} . Consider first the case when \mathbf{j} vanishes: $\partial H_i / \partial x_j$ then has principal axes which are orthogonal. Let these be taken as Cartesian axes, with the neutral point as origin. Then the field at a point on one of these axes has the direction of that axis. Also, since $\text{div } \mathbf{H} = 0$, the diagonal components of $\partial H_i / \partial x_j$ cannot all have the same sign. Let $\partial H_1 / \partial x_1$ and $\partial H_2 / \partial x_2$ have opposite signs. Now consider the field, when there is a current in the z -direction. The direction of the field at points in the

(x, y) plane lies in the (x, y) plane. The lines of force are shown in fig. 1 (a), for the case when the direction of the field belonging to the current is clockwise. The principal axes are no longer perpendicular. The direction of the electromagnetic force is shown in fig. 1 (b) and the gas must flow in the same general direction; the gas will be stretched in the vertical direction in fig. 1. Since the lines of force are frozen into the gas, the principal axes will rotate towards each other. This suggests that the current density will be increased, in which case the situation is unstable, because a small current density will cause a motion which will in turn increase the current density. The current density will then grow until it reaches the proportions of a discharge. The approximation that the lines of force are frozen into the gas then breaks down near the neutral point because the electric field required to drive the current becomes important. A steady state will be reached when the decay of the current density due to the contribution of this accelerating field to $\text{curl } \mathbf{E}$ balances the growth of the current density due to the motion.

Fig. 1



It remains to show that the current density does grow to the proportions of a discharge. In § 5 this result is proved omitting the effect of the pressure gradient; then, if it can be shown that the pressure gradient reinforces the electromagnetic force in the neighbourhood of the neutral point, the result holds *a fortiori*. In § 6 two-dimensional models are considered including the effect of the pressure gradient and it is shown that the condition of mechanical equilibrium requires the current density at an X-type neutral point to be infinite.

This conclusion that the current density will grow near a neutral point but not elsewhere, which appears to draw an absolute distinction between different points in space, may be further clarified if expressed in the following way. If there is a small disturbance in a region of non-zero magnetic field the current density will not become large in this region, but the disturbance will spread in the form of Alfvén waves ; if, however, there is a neutral point anywhere a small disturbance can cause the current density to become large in the neighbourhood of the neutral point.

Fig. 2

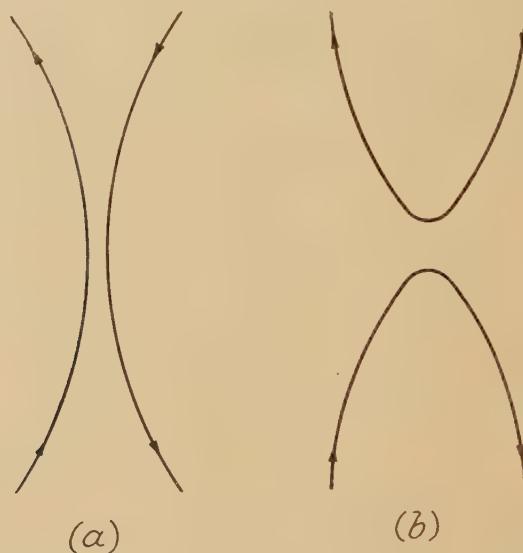
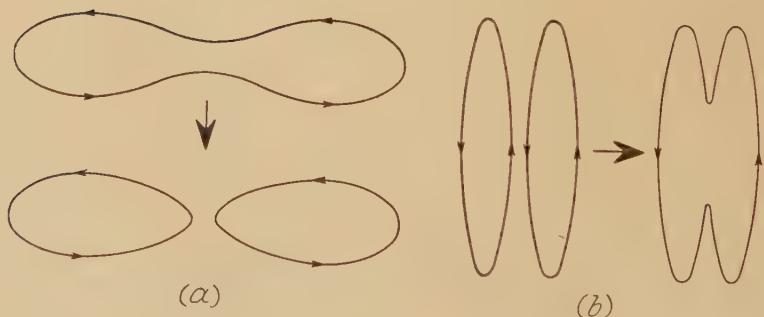


Fig. 3



The contribution of the accelerating electric field to $\partial \mathbf{H} / \partial t$ can be pictured in terms of lines of force. If there are two lines of force as shown in fig. 2 (a), the direction of the current corresponds to a field in the clockwise direction. The field therefore decays in that direction. The

lines of force in fig. 2 (a) can be regarded as being broken and rejoined to form those shown in fig. 2 (b). The total length of the lines of force decreases in the process, and it follows that the energy of the field decreases. This is necessary, since the energy for the discharge must be supplied by the field, if the material is initially static. Figure 3 shows two simple examples. In (a) two parts of a loop of force are close together with their fields in opposite directions, and the result is that the loop of force breaks into two loops, whose total length is less than that of the original loop. In (b) the reverse process occurs, but the length of the final loop is less than the combined length of the original two loops. In both cases field energy is released and field energy from a relatively large region is concentrated on the particles in the neighbourhood of the neutral point.

Because the argument in this section is based on diagrams, we ought to consider the nature of the field in a plane parallel to the paper, but a short distance away from it. At a short enough distance the field is similar to that in the plane of the paper, but there is a small component perpendicular to this plane. The motion of the gas is also similar and $|\text{curl } \mathbf{H}|$ is very large in a small region near the neutral point. The field is not frozen into the gas in this region and the lines of force can be regarded as being broken and rejoined in the way just described. The discharge extends in the direction perpendicular to the paper up to a distance where the change in field is considerable.

§ 5. MATHEMATICAL TREATMENT OF NEUTRAL POINTS NEGLECTING THE PRESSURE GRADIENT

A mathematical treatment of this instability is possible, if the pressure gradient is omitted, and provides an estimate of the time required for the discharge to start. The equations of motion are (3) and the hydrodynamical equation

$$\partial \mathbf{u} / \partial t = -(\mathbf{u} \cdot \nabla) \mathbf{u} + (\text{curl } \mathbf{H}) \wedge \mathbf{H} / 4\pi\mu. \quad \dots \quad (4)$$

We use a frame in which the neutral point is initially a stagnation point of the motion, and then the neutral point remains a neutral point and stagnation point throughout the motion.

Writing u_{ij} and H_{ij} for the tensors $\partial u_i / \partial x_j$ and $\partial H_i / \partial x_j$ and remembering that \mathbf{u} and \mathbf{H} vanish at the neutral point, we obtain

$$\partial H_{ij} / \partial t = -u_{kj} H_{ik} + H_{kj} u_{ik} - H_{ij} u_{kk} \quad \dots \quad (5)$$

and

$$\partial u_{ij} / \partial t = -u_{kj} u_{ik} + (H_{ii} - H_{ii}) H_{ij} / 4\pi\mu. \quad \dots \quad (6)$$

Also

$$\partial \mu / \partial t = -\mu u_{kk}. \quad \dots \quad (7)$$

Equations (5), (6) and (7) determine the time derivatives of u_{ij} , H_{ij} and μ at the neutral point in terms of these variables themselves. If the pressure gradient were included in eqn. (4) higher derivatives of the velocity would be involved and the number of variables would be infinite.

However, it is useful to study these equations and rely on physical arguments to discuss the effect of the pressure gradient, which depends on the state of the rest of the system. Also, the equations corresponding to (5) and (6) at any point other than a neutral point involve higher derivatives of the velocity and magnetic field, so that the mathematical method used here breaks down, and it is necessary to fall back on the physical argument, which has been given in § 3.

We take the case in which all components of both H_{ij} and u_{ij} with either suffix equal to 3, except H_{33} , vanish initially; then they vanish throughout. Typical equations for the other components are

$$\left. \begin{aligned} \partial H_{11}/\partial t &= u_{12}H_{21} - u_{21}H_{12} - H_{11}(u_{11} + u_{22}), \\ \partial H_{12}/\partial t &= u_{12}(H_{22} - H_{11}) - 2u_{22}H_{12}, \\ \partial u_{11}/\partial t &= -u_{11}^2 - u_{12}u_{21} + (H_{12} - H_{21})H_{21}/4\pi\mu, \\ \partial u_{12}/\partial t &= -u_{12}(u_{11} + u_{22}) + (H_{12} - H_{21})H_{22}/4\pi\mu. \end{aligned} \right\} \quad \dots \quad (8)$$

Consider the state in which the current density vanishes; let the axes be chosen so that H_{12} and H_{21} vanish, and let H_{11} be positive and H_{22} negative. Also let all components of u_{ij} vanish, and consider a perturbation in H_{12} , H_{21} , u_{12} or u_{21} . Remembering that μ is always positive, the eqns. (8) show that the signs of the components of H_{ij} and u_{ij} will at first be given by one of the schemes in table 1.

Table 1

H_{11}	H_{22}	H_{12}	H_{21}	u_{11}	u_{22}	u_{12}	u_{21}
+	-	-	+	-	-	+	+
+	-	+	-	-	-	-	-

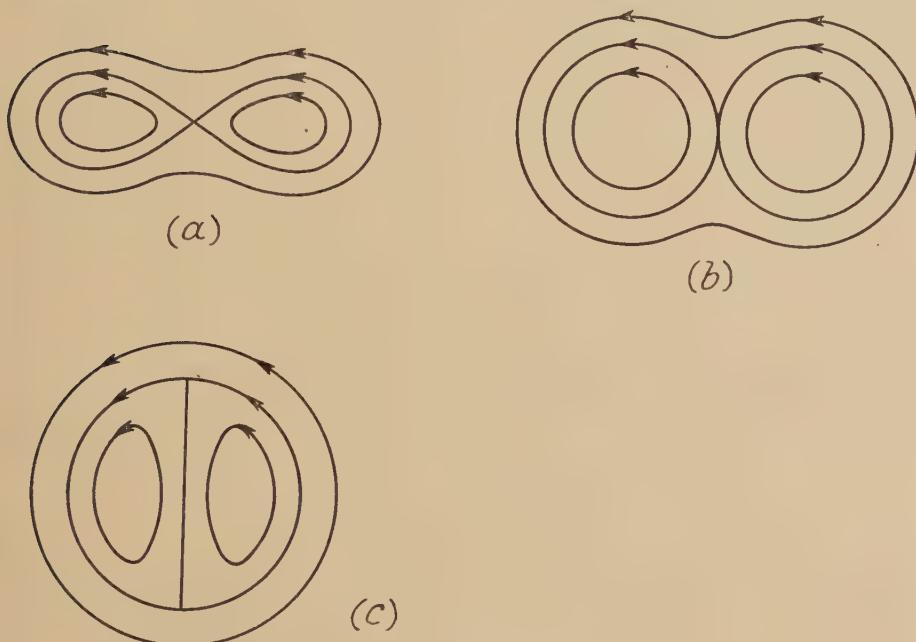
It can then be seen that every term in the derivative of each component has the same sign as that component. Consequently all the components grow in magnitude and, since H_{12} and H_{21} have opposite signs, the current density grows. Until H_{12} , H_{21} , $(4\pi\mu)^{1/2}u_{12}$ and $(4\pi\mu)^{1/2}u_{21}$ are comparable with H_{11} and H_{22} , they grow exponentially with a time constant $(4\pi\mu)^{1/2}/|H_{11} - H_{22}|$. An example has been computed on the EDSAC in the Mathematical Laboratory, Cambridge, by S. Gill. After the components H_{12} , H_{21} , $(4\pi\mu)^{1/2}u_{12}$ and $(4\pi\mu)^{1/2}u_{21}$ become comparable with H_{11} and H_{22} , all the components become infinite in a time of the order of the initial value of $(4\pi\mu)^{1/2}/H_{11}$, as is to be expected from the quadratic form of (8); μ does not increase appreciably until the later stage. We conclude that, if the pressure gradient does not oppose the motion, the situation is unstable in such a way that a discharge can be started by a small perturbation, and that the time required for the current density to grow is not many times larger than the initial value of $(4\pi\mu)^{1/2}/H_{11}$.

§ 6. TWO-DIMENSIONAL MODELS WITH NEUTRAL POINTS

For the purpose of obtaining a simple illustrative model one obvious simplification is to make the field two-dimensional, by taking $H_z = 0$, $\partial H_x/\partial z = \partial H_y/\partial z = 0$. Consider the configuration in static equilibrium.

In the simplest case the lines of force are concentric circles but there is then no X-type neutral point. If there is an X-type neutral point there is a line of force shaped like a figure 8 as shown in fig. 4 (a), where there is zero current density at the X-type neutral point. The shape of the lines of force in equilibrium depends on the relative strength of the field in different regions and can be discussed, using the fact that the energy in any thin tube of force increases with the length of the tube.

Fig. 4



If the field in any particular tube were much stronger than in any of the others, this tube would take up a nearly circular shape, just as in the simplest case all the tubes are circular. If the magnetic energy inside the loops greatly exceeds that outside, the lines of force inside will approximate to concentric circles and in the extreme case the figure 8 will consist of two circles in contact as in fig. 4 (b). If the magnetic energy outside the figure 8 is much the greater, the lines of force outside will approximate to circles and in the extreme case the figure 8 will consist of two D's back to back as in fig. 4 (c). In any intermediate case the configuration will be intermediate between figs. 4 (b) and (c), as shown in fig. 5. In each of these cases the angle between the limiting lines of force at the neutral point is zero. Also the field is symmetrical about the line through the three neutral points, which we now take as x -axis, the origin O being taken at the X-type neutral point. Consider a line of force inside one of the loops of the figure 8 and cutting the x -axis in P and Q, and let its curvature at P

be K_P and at Q , K_Q . Obviously for fig. 4 (b) $|K_P|=|K_Q|$ and for fig. 4 (c) $|K_P|>|K_Q|$ so that in any case

$$|K_P|\geq|K_Q|. \quad \dots \dots \dots \quad (9)$$

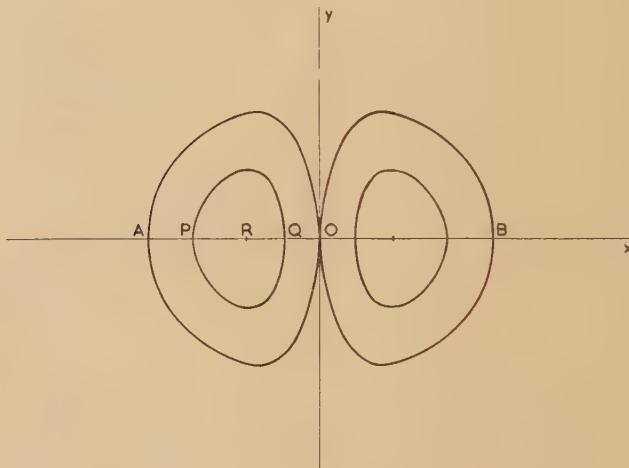
This result can be used to prove that $|\operatorname{curl} \mathbf{H}|$ is infinite at O if $|\mathbf{H}|$ is finite at A and B .

The equation for static equilibrium is $\mathbf{j} \wedge \mathbf{H}/c = \nabla p$ which yields $\operatorname{curl}(\mathbf{j} \wedge \mathbf{H})=0$ and for two dimensional fields this reduces to

$$(\mathbf{H} \cdot \nabla) \mathbf{j} = 0. \quad \dots \dots \dots \quad (10)$$

Equation (10) shows that \mathbf{j} is constant on a line of force, but (10) is automatically satisfied at a neutral point, so that \mathbf{j} can have any value at a neutral point. The vector potential \mathbf{A} for the magnetic field can be taken

Fig. 5



as $(0, 0, A)$ and then \mathbf{j} is $(0, 0, -c\nabla^2 A/4\pi)$. On a line of force A is constant and (10) shows that $\nabla^2 A$ is also constant, so that at a pair of points situated like P and Q

$$\left. \begin{aligned} A_P &= A_Q, \\ (\nabla^2 A)_P &= (\nabla^2 A)_Q. \end{aligned} \right\} \quad \dots \dots \dots \quad (11)$$

Consider the variation of A on the x -axis, so that $\partial^2 A / \partial x^2$ can be written

$$\frac{dA}{dx} \frac{d}{dA} \left(\frac{dA}{dx} \right),$$

and $\partial^2 A / \partial y^2 = K dA / dx$ where K is the curvature of the line of force defined to be positive when the line of force is convex towards the direction of positive x .

Combining these results

$$\frac{1}{2} \frac{d}{dA} \left[\left(\frac{dA}{dx} \right)^2 \right] + K \frac{dA}{dx} = \nabla^2 A. \quad \dots \dots \dots \quad (12)$$

The points P and Q move along the x -axis as A varies and (11) and (12) yield

$$\frac{d}{dA} \left[\left(\frac{dA}{dx} \right)_Q^2 - \left(\frac{dA}{dx} \right)_P^2 \right] = 2 \left(K_P \left(\frac{dA}{dx} \right)_P - K_Q \left(\frac{dA}{dx} \right)_Q \right). \quad \dots \quad (13)$$

Now $K_P < 0$ and $K_Q > 0$. Suppose that P and Q move apart as A increases so that $(dA/dx)_P < 0$ and $(dA/dx)_Q > 0$. Then the right-hand side of (13) ≥ 0 as

$$|dA/dx|_Q \leq |K_P/K_Q| |dA/dx|_P.$$

Since $(dA/dx)_P$ and $(dA/dx)_Q$ tend to zero as P and Q approach R, and remembering (9), we conclude that

$$|dA/dx|_Q > |dA/dx|_P \text{ or } |\mathbf{H}_Q| > |\mathbf{H}_P|.$$

The above argument is valid so long as PQ lies inside AO and hence if $|\mathbf{H}|$ is not zero in the neighbourhood of A, $|\mathbf{H}|$ is finite at any point $(-\epsilon, 0)$ where ϵ is small but not zero. Similarly if $|\mathbf{H}|$ is not zero in the neighbourhood of B, $|\mathbf{H}|$ at $(+\epsilon, 0)$ is finite. The field is directed in opposite directions at $(-\epsilon, 0)$ and $(+\epsilon, 0)$, hence at O the field must change discontinuously and $\text{curl } \mathbf{H}$ must be infinite.

The situation in which there is an infinite current density may be regarded as the extreme case of constriction. Constriction is usually discussed in connection with a field whose lines of force are concentric circles, and then the constriction is usually limited by the gas pressure. The difference between this case and that of an X-type neutral point is that in the latter the material can escape from the region of high current without crossing lines of force.

§ 7. ORBITS OF THE PARTICLES

The foregoing arguments for particular conditions show that the current density at a neutral point increases so long as eqn. (3) is a valid approximation, and we are justified in stating that the current density becomes very large. A thorough discussion of the behaviour of such a discharge when our approximation breaks down will not be given here, as the calculation of the current due to the accelerated particles, when they have left the accelerating region, is too difficult and depends on the configuration of the field outside the accelerating region. A rough discussion is attempted in order to obtain an estimate of the importance of these discharges.

It has been seen that the accelerating region must be exceedingly thin in one direction, which is clearly the horizontal direction in fig. 1. It may extend to any distance in the vertical direction and so can be regarded as a very thin sheet. The effect of the magnetic field on the orbit of a particle during acceleration is important. Consider an orbit passing through the neutral point in a direction almost perpendicular to the plane of fig. 1. If it deviates in the horizontal direction it is brought back by the magnetic field, so that the orbit stays in the accelerating region even though this is very thin. If it deviates in the vertical direction, the magnetic field bends

it further away, so that the orbits fan out in the vertical direction. After leaving the accelerating region the particles can be regarded as moving along lines of force, if these are regarded as moving with the material. Now in the plane containing the perpendicular to the plane of fig. 1 and one of the other principle axes, the lines of force all pass through the neutral point. Consequently an orbit approximately follows one of these lines of force, and since the orbits in the accelerating region fan out, they will continue to spread after leaving the accelerating region, and the current density due to the accelerated particles will be much smaller than it is in the accelerating region. Outside the accelerating region there is a background of unaccelerated particles, and, since the accelerated particles will be considerably less numerous than these, they will probably neutralize the space charge and current density of the accelerated particles. We therefore suppose that a steady state is set up as described in § 4.

§ 8. APPLICATIONS

It is now desirable to obtain a rough estimate of the voltage driving a discharge at a neutral point. Suppose that a steady state is set up as described in § 4. The electric field can be considered as the sum of the part driving the discharge and the induced field $-\mathbf{u} \wedge \mathbf{H}/c$. Then since $\text{curl } \mathbf{E}$ must vanish in a steady state, it can be concluded that the electric field driving the discharge is of the same order as the induced field outside the discharge. In fig. 1 the induced field is everywhere directed into the paper and so also is the field required to drive the discharge. If the particles are accelerated over a distance l , they will acquire energies of order $e|\mathbf{u}||\mathbf{H}|l/c$. In the following it will be assumed that the discharge extends in the direction of the current over a distance of order a , the characteristic length of the system. It is found that collisions of the accelerated particles are not important in the applications discussed and then $l \sim a$. Particles with momentum less than $e|\mathbf{H}|a/c$ move in orbits which spiral round the lines of force. For particles with relativistic energy the corresponding energy is approximately $e|\mathbf{H}|a$, so that particles accelerated at a neutral point do not acquire sufficient momentum to escape across the magnetic field. The energies involved are nevertheless very large; rough values for particular applications will now be briefly discussed.

The values of the relevant quantities are most accurately known for Hoyle's suggested theory of the aurora (Hoyle 1949). According to this a beam of ionized gas with a magnetic field frozen into it is emitted by the sun, neutral points occur in the neighbourhood of the earth, and the aurora is due to particles accelerated at these neutral points, which then travel along lines of force until they penetrate the atmosphere of the earth. The motion of the beam sets up currents at the neutral points which flow in a particular direction; the pressure gradients set up by the motion of the beam then reinforce the electromagnetic forces near the neutral point, and the result obtained in § 5 shows that discharges will occur. When a

steady state is set up, the neutral points are stationary relative to the earth. The velocity of the beam is inferred from the delay between the observation of a solar flare and the commencement of a magnetic storm and is about 10^8 cm/sec. Hoyle estimates the strength of the magnetic field in the beam at about 10^{-3} gauss, so that the electric field is of order 10^{-3} volts/cm. The only collisions that could be important are collisions with charged particles (i.e. encounters in which the particle suffers a large deflection) and the mean free path for such collisions is of order w^2/ne^4 , where w is the energy of the particle. n is probably about 100 particles/cm³ and even if w is only the thermal energy, say 10^{-14} ergs., the mean free path is of order 2×10^7 cm. These collisions can therefore be neglected, because the mean free path increases as the energy of the particle increases. Hoyle supposes that the particles are accelerated over a distance 4×10^7 cm. This may be an underestimate, but it is sufficient to obtain particles of energy 4×10^4 ev, and this is the energy required for the particles to penetrate to a height of 100 km above the surface of the earth.

Giovanelli (1947, 1948) first suggested the possibility of discharges occurring at neutral points in connection with solar flares, which occur in the neighbourhood of large sunspots. He discusses collisions with neutral atoms and finds that these will be unimportant, if $|\mathbf{E}|$ exceeds E_L , where E_L depends on the height in the chromosphere, and has its maximum value, about 10^{-3} e.s.u., at the base of the chromosphere. If $|\mathbf{H}| \sim 1000$ gauss, this would require only $|\mathbf{u}| \sim 3 \times 10^4$ cm/sec so that collisions can again be neglected. If $4\pi\mu |\mathbf{u}|^2 \sim |\mathbf{H}|^2$, and $\mu \sim 10^{-13}$ g/cm³, $|\mathbf{u}|$ must be of order 10^9 cm/sec. (This estimate is made purely on theoretical grounds and no such large velocities have been observed.) Then taking $a \sim 10^9$ cm, particles will be accelerated to 10^{13} ev. Even if this estimate is a factor of 1000 too high, soft cosmic rays would be produced and they could account for the large increases sometimes observed in the total intensity of cosmic rays at the time of intense solar flares (Forbush 1946, Neher and Roesch 1948).

We may also speculate on the possibility of discharges at neutral points in interstellar space using the values given by Fermi (1949). He describes a process by which particles could acquire energy over a very long time, and which shows promise of explaining the spectrum of cosmic rays. He discusses the collision processes which occur and they can certainly be neglected in a discharge. He gives the values $|\mathbf{H}| \sim 10^{-5}$ gauss, $|\mathbf{u}| \sim 3 \times 10^6$ cm/sec and $a \sim 10^{19}$ cm, which lead to an energy of 3×10^{12} ev. These values are very uncertain; Batchelor (1950) believes the value of the magnetic field strength to be considerably too large. It may also be noted that Fermi requires heavy positive ions to be accelerated to about 10^9 ev before his acceleration process will work, and that this could be achieved in solar flares and of course in other stars.

Electrical discharges are probably important as sources of radio noise. Outbursts of radio noise are known to be associated with solar flares, and it is possible that radio stars are also associated with discharges.

ACKNOWLEDGMENTS

The author wishes to thank Mr. F. Hoyle for introducing him to this subject and for valuable discussions and Dr. G. R. Giovanelli for valuable discussions. He is also grateful to Professor S. Chapman for advice, which led to the investigation contained in § 6.

REFERENCES

ALFVÉN, H., 1950, *Cosmical Electrodynamics* (Oxford).
BATCHELOR, G. K., 1950, *Proc. Roy. Soc. A*, **201**, 405.
COWLING, T. G., 1945, *Proc. Roy. Soc. A*, **183**, 453.
DUNGEY, J. W., 1950, *Proc. Cam. Phil. Soc.*, **46**, 651.
ELSASSER, W. M., 1947, *Phys. Rev.*, **72**, 821.
FERMI, E., 1949, *Phys. Rev.*, **75**, 1169.
FORBUSH, S. E., 1946, *Phys. Rev.*, **70**, 771.
GIOVANELLI, R. G., 1947, *Mon. Not. R. Astr. Soc.*, **107**, 338; 1948, *Ibid.*, **108**, 163.
HOYLE, F., 1949, *Some Recent Researches in Solar Physics* (Cambridge), p. 103.
NEHER, H. V., and ROESCH, W. C., 1948, *Rev. Mod. Phys.*, **20**, 350.
SWEET, P. A., 1949, *Mon. Not. R. Astr. Soc.*, **109**, 507.
WALÉN, C., 1949, *Ark. f. Mat. Astr., och Fys.*, **33A**, No. 18, pp. 19–24.

LXXVII. *The Beta-Gamma Angular Correlation of ^{76}As*

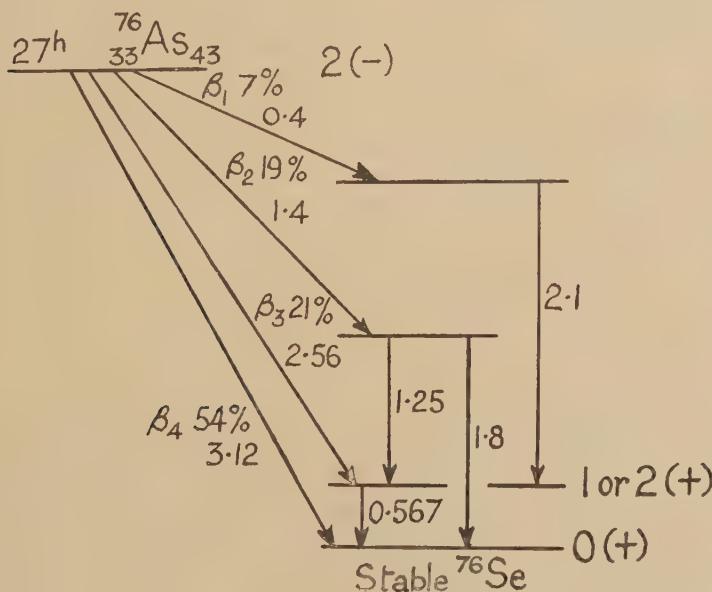
By H. Rose
Cavendish Laboratory, Cambridge *

[Received March 22, 1953]

ABSTRACT

The beta-gamma angular correlation of ^{76}As has been measured at specific beta-ray energies using a thin magnetic lens beta-ray spectrometer. The experimental results cannot be explained by a single nuclear matrix element if the gamma-ray is considered as electric quadrupole radiation, but if the gamma-ray is assumed to be magnetic dipole radiation good agreement with the theoretical predictions is obtained for either the first forbidden matrix element $\int \sigma \times r$ of the axial vector interaction alone or the corresponding $\int \beta \sigma \times r$ of the tensor interaction alone.

RIDGWAY AND PIPKIN (1952) have measured the beta-gamma angular correlation of ^{76}As which occurs for beta-ray spectrum 3 (see figure) in



Decay scheme of ^{76}As (Marty *et al.* 1949).

coincidence with the 0.567 Mev gamma-ray. Regarding the latter as electric quadrupole radiation they concluded that it would be necessary to consider a mixture of nuclear matrix elements to explain the observed

* Communicated by Mr. E. S. Shire.

correlation. We have re-measured this angular correlation as a function of beta-ray energy and have obtained results which confirm those of Ridgway and Pipkin. However, we show below that it is possible to interpret the experimental results in terms of a single matrix element if the 0.567 mev gamma is assumed to be magnetic dipole radiation.

The experiment was carried out with the thin magnetic lens beta-ray spectrometer previously used (Rose 1952) but with improved magnetic shielding of the photomultipliers from the spectrometer lens coil. The ^{76}As sources were prepared by first dissolving 'Specpure' As_2O_3 of high specific activity (about 0.7 mc mg^{-1}) in concentrated ammonia. Some of the solution was then placed on to thin aluminium foil (0.25 mg cm^{-2}) and evaporated down to form sources whose thickness was always below 1 mg cm^{-2} . A 1.7 g cm^{-2} lead absorber was placed in front of the NaI crystal of the gamma-detector to exclude Compton scattered quanta. In other respects the experiment was performed in a manner similar to that employed for previous investigations (Rose 1952).

The values obtained for the differential correlation coefficient a (assuming the angular correlation function $W(\theta) = 1 + a \cos^2 \theta$), measured at three beta-ray energies, are listed in the table. No measurements of the correlation were made below 1.4 mev since at this energy spectrum 2 begins to contribute to the measured beta-gamma coincidence rate.

Energy of beta-rays (mev)	a (experimental)	a (theoretical)	
		$\int \sigma \times \mathbf{r}$ (axial vector)	$\int \beta \sigma \times \mathbf{r}$ (tensor)
1.42	$+0.056 \pm 0.018$	+0.041	+0.066
1.69	$+0.065 \pm 0.020$	+0.065	+0.071
2.00	$+0.076 \pm 0.024$	+0.075	+0.075

^{76}Se is an even-even nucleus and therefore its ground state has zero spin and even parity. Tomlinson and Ridgway (1952) have found that spectrum 4 has a curved Kurie plot which can be fitted by the correction factor $p^2 + q^2$ (Konopinski 1943), indicating that ^{76}As has spin 2 and odd parity. They also measured the K -conversion coefficient of the 0.567 mev gamma-ray assuming the decay scheme of Siegbahn (1947) and obtained agreement with the theoretical prediction for electric quadrupole radiation, thereby assigning spin 2 and even parity to the first excited state of ^{76}Se . With these spin assignments it is not possible to get agreement between the observed correlation and the theoretical predictions of Falkoff and Uhlenbeck (1950) assuming a single first forbidden nuclear matrix element operative in spectrum 3, as has been pointed out by Ridgway and Pipkin (1952).

Since Goldhaber and Sunyar (1951) have shown that the first excited state of even-even nuclei is usually $2(+)$, it is perhaps improbable that the 0.567 mev transition can be magnetic dipole. Nevertheless, it is interesting to note that a fit with the theory for a single matrix element is possible assuming that the first excited state of ^{76}Se is $1(+)$. Then either the first forbidden matrix element $\int \sigma \times \mathbf{r}$ of the axial vector interaction alone or the corresponding $\int \beta \sigma \times \mathbf{r}$ of the tensor interaction alone can account for the correlation, as indicated in the table. Moreover the tables of Rose *et al.* (1951) yield for the 0.567 mev gamma the values $\alpha_K = 1.8 \times 10^{-3}$ for electric quadrupole and $\alpha_K = 1.3 \times 10^{-3}$ for magnetic dipole radiation. In view of the uncertainty of Siegbahn's decay scheme (10–15% in gamma-ray intensity) it is perhaps doubtful whether Tomlinson and Ridgway's value for the K -conversion coefficient of $(2.0 \pm 0.2) \times 10^{-3}$ can distinguish uniquely between these two types of transition.

I wish to express my thanks to Mr. N. Sutin for assistance in the preparation of the sources.

REFERENCES

FALKOFF, D. L., and UHLENBECK, G. E., 1950, *Phys. Rev.*, **79**, 334.
 GOLDHABER, M., and SUNYAR, A. W., 1951, *Phys. Rev.*, **83**, 906.
 KONOPINSKI, E. J., 1943, *Rev. Mod. Phys.*, **15**, 209.
 MARTY, N., LABEYRIGUE, J., and LANGEVIN, H., 1949, *Comptes Rendus*, **288**, 1722.
 RIDGWAY, S. L., and PIPKIN, F. M., 1952, *Phys. Rev.*, **87**, 202.
 ROSE, H., 1952, *Phil. Mag.*, **43**, 1146.
 ROSE, M. E., GOERTZEL, G. H., and PERRY, C. L., 1951, ORNL-1023.
 SIEGBAHN, K., 1947, *Arkiv. Mat. Astron. Fysik*, **34A**, No. 7.
 TOMLINSON, E. P., and RIDGWAY, S. L., 1952, *Phys. Rev.*, **88**, 170.

**LXXVIII. A Theory of Work-Hardening of Metals
II : Flow Without Slip-Lines, Recovery and Creep**

By N. F. MOTT

H. H. Wills Physical Laboratory, University of Bristol*

[Received April 15, 1953]

SUMMARY

The author's previous theory of work-hardening is extended to account for fine slip. It is suggested that fine slip and coarse slip both have their origin in Frank-Read sources, but that fine slip occurs when the dislocations move slowly. This occurs when the stress Gb/l required to obtain dislocations from a source is less than the stress required to drive them through the obstacles in the lattice without the help of thermal activation. The initial stages of deformation are normally by fine slip, and the hardening in this region is shown to be much slower than for coarse slip. In creep the deformation is normally by fine slip; a discussion is given on this basis of logarithmic creep, which agrees better with experiment than previous exhaustion theories. A new theory is given of Andrade's β -creep, which relates it closely to steady-state creep. Finally a discussion is given of the formation of vacancies during creep.

§ 1. INTRODUCTION

IN a previous paper (Mott 1952 a) the present author put forward a theory of work-hardening of face-centred cubic metals, applicable primarily to slip at comparatively low temperatures. The main points of the theory were :

(a) Slip has its origin at Frank-Read sources.

(b) Once a source has started to generate dislocations, it will continue to do so until the stress at the source due to these dislocations is of the order Gb/l , the stress required to operate the source. The former stress may be due to moving dislocations (Fisher, Hart and Pry 1952) or to dislocations piled up against a distant barrier. About 1000 dislocation rings are normally formed before the source stops. This process we call 'coarse slip'. It was explained by assuming that dislocations can move with a speed that is a significant fraction (say one half) of the speed of sound; the momentum of a moving dislocation then keeps the source going until the local stress in its neighbourhood drops by a considerable factor. Coarse slip is thus to be explained by the 'dynamic' behaviour of dislocations.

* Communicated by the Author.

(c) If there are no barriers (grain boundaries or sessile dislocations), the dislocations pass out of the crystal. There is then little or no hardening. Hardening is mainly the result of the strains round piled-up groups of dislocations which are trapped in the metal. These groups are locked in position by the Lomer-Cottrell mechanism.

The purpose of this paper is to extend the theory in the following ways :

(i) To give a discussion of 'fine slip', that is strain in the form of a large number of very fine slip lines. This we believe to be due to a non-dynamic action of Frank-Read sources, both dynamic and non-dynamic action being possible under suitable conditions.

(ii) To discuss thermal recovery of a cold-worked metal.

(iii) To describe in terms of dislocation movement some of the observed forms of creep, particularly logarithmic creep, Andrade's β -creep and steady-state creep.

(iv) To discuss the rate of self-diffusion in materials undergoing creep.

§ 2. A THEORY OF FINE SLIP

There is much evidence that, as well as extension by well marked slip lines of height $c.$ 2000 Å, or the clusters of such lines observed by Heidenreich and Shockley (1948) and by Brown (1951, 1952), slip can occur by the formation of a much larger number of fine slip lines, the step height being $c.$ 50 Å or less.* Kuhlmann-Wilsdorf *et al.* (1952) (see also Wilsdorf and Kuhlmann-Wilsdorf 1951, 1952) have observed such lines on aluminium crystals and polycrystalline specimens deformed at ordinary temperatures and speeds; according to their observations fine slip may account for about 10% of the total strain. Hanson and Wheeler (1931) first showed that at sufficiently low rates of strain no slip lines appear on polycrystalline aluminium, and according to McLean (1952) strain in creep in aluminium at 200°C is due to fine slip; apparent coarse slip lines are simply due to clustering of fine slip. Then Brown and Honeycombe (1951) have observed fine slip on surfaces that have been electro-polished but not mechanically polished. Smith and Dewhurst (1949) observe that copper containing a fine dispersion of aluminium oxide does not show slip lines, and slip lines in the unoxidized region terminate sharply at the boundary of the oxidized region. Leibfried (1950) shows that the extension of aluminium under creep conditions is not jerky, as it would be if 1000 dislocations spread over any appreciable area in a microsecond. Finally it is usually stated that slip lines are not observed for strains of less than about 2%; Honeycombe (1950) has shown that in aluminium weak deformation bands appear for strains of this order.

* This is not to be confused with flow by migration of vacancies, of which the properties were worked out first by Nabarro (1948) and later by Conyers Herring (1950), and which has been observed at high temperatures and small strains by, for instance, Udin, Shaler and Wulff (1949) for copper.

Coupled with these facts we may add the behaviour of hexagonal crystals, where according to Andrade and Roscoe (1937) weak slip lines first appear and then grow in height, in sharp contrast to the behaviour of cubic crystals where the steps appear having their full height, their number or length but only to a slight extent their height increasing as the strain increases (Chen and Pond 1952).

Apart from the direct observation of Kuhlmann-Wilsdorf *et al.*, it seems to the author that the demonstration by a number of authors (Gough and Wood 1936, 1938, Heidenreich 1951, Hirsch 1952, Warren and Averbach 1952) of the formation of slightly disoriented crystallites in heavily cold-worked metals points strongly to part of the deformation being by fine slip. The formation of boundaries after coarse slip is only possible if dislocations can 'climb' out of piled-up groups (Mott 1951), which is unlikely at room temperature;* fine slip, on the other hand, produces dislocations dispersed on a large number of planes, so that they can line up without climbing out of their slip planes. We suggest that the mechanism is the same as for Honeycombe's deformation bands.

The behaviour of some oxidation-hardened materials can perhaps be explained by the author's previous theory. If the oxide particles greatly cut down the slip distance L , the number of dislocations on each line, given by the formula, in which l is the length of the source

$$n \simeq 2\pi L/l$$

will be greatly reduced too. But the other phenomena do suggest that a Frank-Read source may under suitable circumstances behave in a way quite different from that described previously.

Now it was pointed out in the author's previous paper that, if there is some frictional force which prevents dislocations from acquiring a speed anywhere near that of sound, then, as soon as a source had produced one dislocation ring, it would stop acting, because of the stress at the source due to the dislocation ring itself; it would not have momentum enough to keep it going. The source could not produce further rings until the applied stress was increased. Thus all available sources would be quickly brought into play, none of them would give many dislocations, and the slip on each active plane would increase with stress. We suggest that this is just what is happening when fine slip is observed, and in hexagonal crystals.

We have now to enquire under what conditions dynamic and non-dynamic movement of dislocations will occur. It is certainly possible that there will be a transition from dynamic to non-dynamic motion as the stress is lowered or the temperature raised. The calculations

* It has been suggested (Mott 1952 a, Cottrell 1952) that these crystallites are formed by polygonization due to 'climb' made possible by the vacancies formed by moving dislocations. However Gay and Kelly (1953) have shown that the crystallites are formed in nickel cold-worked at room temperatures, at which little movement of vacancies is to be expected; so this cannot be the complete explanation.

of Nabarro (1951) on damping of a dislocation moving through an otherwise perfect lattice make this prediction, though it has not yet proved possible to say at what stress or temperature the transition should occur. It is possible that fine slip may be due to damping of this kind. The observations listed above suggest, however, another explanation. We propose that non-dynamic motion can occur if the slip-planes contain obstacles to the motion of dislocations, each obstacle extending in the plane over a few atomic distances only.* A moving dislocation may be hung up by these obstacles, much as illustrated in fig. 2. We introduce the stress σ_0 required to force a dislocation through these obstacles (cf. § 3). If the applied stress σ is greater than σ_0 , dynamic motion is to be expected. If, on the other hand, σ is less than σ_0 , it is suggested that the dislocation will frequently be held up by the obstacles, and released by temperature, and then held up by the next obstacle, and so on. It moves forward in a series of little jerks. The result on the movement will be equivalent to a frictional force.

The stress required to operate a Frank-Read source in an otherwise perfect lattice is defined by

$$\sigma_{FR} = Gb/l,$$

where l is the length of the source. If for a given source σ_{FR} is greater than σ_0 , then dynamic generation of dislocations is to be expected. If, however, sources exist for which σ_{FR} is less than σ_0 , the dislocation rings from the source will not acquire the speed of sound, and fine slip is to be expected.

Let us now see what these obstacles may be, and how the facts listed above may be explained. The possible application to materials containing dispersed oxide is obvious, and will not be discussed further. In pure metals we suggest, following Cottrell (1952), that they are screw dislocations which cut the slip plane. If these are a mean distance l_0 apart, we shall show in § 3 that

$$\sigma_0 = \alpha Gb/l_0,$$

where, according to calculations by Stroh (1953), α is of order 2 or 3. Since l_0 and the mean value of l for the sources are expected to be of the same order, we may expect that just a few of the longer sources will give fine slip, but most of them will give coarse slip. It is, of course, the sources of large l which are the first to generate dislocations. We thus see why, as the stress is raised, fine slip precedes coarse slip. Also in creep the stress does not normally rise above the value needed to operate these few which have large values of l , and this ensures fine slip only. And, finally, as Hollomon (1952) has pointed out, dislocation lines ending on an electropolished surface may behave like Frank-Read sources of twice the normal length (see also Mott 1952 a). They will thus give fine slip.

* In contradistinction to sessile dislocations, which provide barriers along a line in the slip plane.

We now investigate the strain hardening curve to be expected in a substance with N sources per unit volume each giving non-dynamic slip. If after a strain ϵ each source has generated n dislocations, then with the author's previous notation

$$\epsilon = NLL'bn. \quad \dots \dots \dots \quad (1)$$

The density of groups of edge dislocations is NL' per cm^2 , so that the mean distance r between them is $1/(NL')^{1/2}$. The internal strain is thus given by

$$\sigma_i \sim Gbn/2\pi r \sim Gbn(NL')^{1/2}/2\pi. \quad \dots \dots \dots \quad (2)$$

Eliminating n between (1) and (2) we have

$$\sigma_i/G = \epsilon/2\pi L(NL')^{1/2}. \quad \dots \dots \dots \quad (3)$$

The flow stress with this model is $\sigma = \sigma_{FR} + \sigma_i$. It will be seen that (3) represents a rate of hardening *linear* in the strain.

Now in single crystals and even in polycrystalline metals (French and Hibbard 1950, Hollomon 1945) some region of linear hardening frequently precedes the 'parabolic' hardening (fig. 1). We suggest that this may coincide with the region of fine slip. This hypothesis is in agreement with observations of Crussard (private communication, see also Jaoul and Crussard 1952), who states that in the region of linear hardening there are always bands on the crystal surface where no slip lines have appeared. It also agrees with the observation of Nishimura and Takamura (1952) that with pure aluminium crystals there are, in the initial region of slow hardening, no slip lines but weak deformation bands of the type observed by Honeycombe (1950).

To obtain a model to give the extent of linear hardening, let us arbitrarily assume that slip-band formation throughout the specimen and parabolic hardening set in when the flow stress has doubled, and thus when σ_{FR} and σ_i are equal. This gives

$$Gb/l \sim G\epsilon/2\pi L(NL')^{1/2}.$$

If we take L equal to L' and suppose that slip lines and parabolic hardening begin when $\epsilon \sim 0.01$, this gives

$$N = 10^{-4}l^2/(2\pi)^2b^2L^3.$$

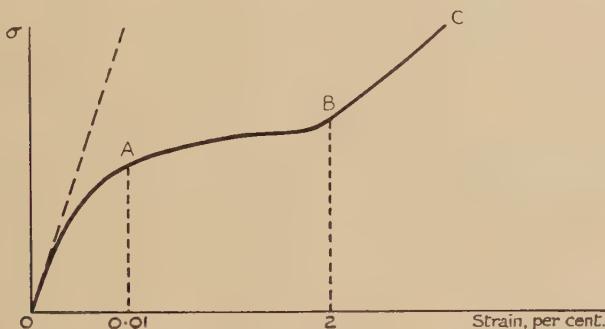
If we take $l \sim 10^{-4} \text{ cm}$, $L \sim 10^{-2} \text{ cm}$, this gives $N \sim 10^8 \text{ cm}^{-3}$, a not unreasonable figure for the number of sources active in fine slip. The number n of dislocations from each source is then according to (1) about 100. It need hardly be said that these figures can be varied within wide limits.

For smaller strains (c. 0.01%) we see that n drops below unit. The above analysis is then no longer applicable. Here we must be in the region of 'exhaustion' hardening (Mott and Nabarro 1948) in which sources of very large l (small σ_{FR}) are used first, and then sources of somewhat larger σ_{FR} . The assumption of constant N will not be a good

approximation in this region. We suppose that the typical stress-strain curve (at low temperatures) of metallic single or polycrystals is somewhat as in fig. 1.

This curve, it should be emphasized, is for conditions in which the dislocations nearly all remain in the metal, for instance for polycrystals. In the phenomenon of easy glide, in which little or no hardening of single crystals was observed for strains up to 30% in Andrade and Henderson's (1951) work—and up to 6% in Lücke and Lange's (1952)—it is possible that the dislocations pass out of the crystal. On the other hand, it is possible that the phenomenon is of the type described above, the slip distance L being particularly large, but still smaller than the dimensions of the crystal. Either explanation would fit the observation of Lücke and Lange, that the phenomenon does not occur when there is slip on more than one set of planes; sessile dislocations would then be formed by the Cottrell-Lomer mechanism.

Fig. 1



Typical stress-strain curve.

OA is the 'exhaustion' region, in which a few sources of particularly large l (small σ_{FR}) are used up; AB is the region in which a large number of sources each generate a few dislocations without dynamic motion; BC is the parabolic region in which coarse slip bands are formed throughout the crystal.

It should be pointed out also that, if an appreciable proportion of the dislocations end up in deformation bands, the hardening will be less than that given by (3).

Lücke and Lange report easy glide only in 'pure' aluminium (99.99%), not in crystals of purity 99.6%. We consider, for reasons put forward in § 9, that any 'Cottrell locking' will greatly decrease the proportion of fine slip—and thus, other things being equal, the range of linear hardening. In a recent paper Blewitt (1953) has obtained linear hardening up to strains of the order 50% for copper crystals of purity 99.999%.

Turning finally to hexagonal crystals, we suggest that in them there may be a high density of immobile dislocations parallel to the hexagonal axis, and that for this reason slip in the basal planes is not dynamic.

We cannot, however, say whether the linear hardening formula (3) is likely to apply to these crystals, as we do not know whether dislocations remain in them, or whether they escape.

§ 3. MOVEMENT OF AN EDGE DISLOCATION THROUGH AN ARRAY OF SCREW DISLOCATIONS

In this section we shall discuss in greater detail Cottrell's model of an edge dislocation moving in a plane which is cut by an array of screw dislocations. At each of these 'crossing points' the dislocation can be held up, as explained in the last section. The motion forward envisaged is as represented in fig. 2. Our aim is to calculate the stress σ_0 required for dynamic motion.

We suppose, as before, that the number of crossing points per unit area of slip plane is $1/l_0^2$. Each time that the edge dislocation passes such a point a jog is formed, for which we write the energy αGb^3 ; α is a numerical factor which Stroh (1953 b) estimates as 2 or 3 (in close-packed structures). Following Cottrell (1952), however, we suppose that in the presence of a stress σ the activation energy for the production of a jog is reduced to

$$U = \alpha Gb^3 - \sigma l_0 b^2: \dots \dots \dots \quad (4)$$

The proof is as follows: we take the energy per unit length of a dislocation to be $\frac{1}{2}Gb^2$ (Frank 1950). Then the form of a dislocation at rest acted on by a stress σ is given by the equation

$$\frac{1}{2}Gb^2 d^2y/dx^2 = \sigma b,$$

whence

$$y = (\sigma/G)x^2/b.$$

If a dislocation line is anchored at two points A, B such that AB is perpendicular to the stress (fig. 2), its inclination at either extremity to the line AB is

$$dy/dx = \sigma l_0/Gb,$$

where l_0 is the length AB. The total force on the locking points, A, B, $Gb^2 dy/dx$, is thus $\sigma l_0 b$, and the work done if the dislocation moves forward through one Burgers vector is $\sigma l_0 b^2$. Formula (4) follows. Thus, if the stress is greater than

$$\sigma_Y = \alpha Gb/l_0,$$

the dislocation cannot be held up by the obstacles. If $\sigma < \sigma_Y$, on the other hand, it can be held up; but we cannot necessarily identify σ_Y with σ_0 , our critical velocity for dynamic motion, because, even if σ is less than σ_Y , a dislocation, once it has broken away from an obstacle, may acquire sufficient kinetic energy to break through the next one.

Figure 2 shows the way in which a dislocation may be expected to break away from an obstacle A. A succession of jumps of this type will bring the whole dislocation forward through a distance l_0 . It will be seen that the force pulling the dislocation past A is much less than that

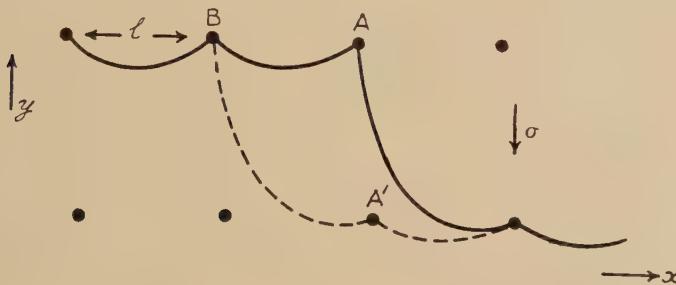
pulling it past the next position A'. The question is, will the kinetic energy acquired in going from the full position to that shown by the dotted line compensate for this, so that the dislocation will break past A' too. We have not been able to answer this with any certainty, but a calculation given in the appendix suggests that it will not. We therefore tentatively equate σ_0 to σ_Y .

When σ is less than σ_0 and a dislocation moves forward in a series of jerks, we shall take its mean velocity to be

$$\nu l \exp(-U/kT),$$

where ν is an atomic frequency of order 10^{12} . Apart from fine slip, we shall use motion of this sort to account for logarithmic creep. Doubtless also this sort of impeded motion of free dislocations accounts for the hysteresis loop of single crystals under reversed stress and for the Bauschinger effect in single crystals.

Fig. 2



Showing the way in which a dislocation line moves forward in jerks, from the position shown by the full line to that shown by the dotted. The dots represent the intersections of screw dislocations with the plane of slip.

§ 4. LOGARITHMIC CREEP

We shall now apply the concepts of the last two sections to a discussion of creep.

For transient creep two formulae have been proposed ; that of Andrade, which relates the strain ϵ to the time t according to the relation

$$\epsilon = \beta t^{1/3}, \quad \dots \dots \dots \dots \dots \dots \quad (5)$$

and logarithmic formulae of the type

$$\epsilon = \alpha \{\log(\gamma t + 1)\}^s, \quad \dots \dots \dots \dots \dots \dots \quad (6)$$

where s is a constant. Andrade's formula is valid for metals and indeed other substances over a wide range of temperatures ; as far as we know no theoretical deviation has been given. One will be attempted in § 6 of this paper. The logarithmic formula, with s equal to $\frac{2}{3}$, was proposed for alloys by Mott and Nabarro (1948) and based on an 'exhaustion' theory ; it was shown to be in agreement with experiment for certain alloys at

low temperatures by Davis and Thompson (1950), but with much too large a value of γ . Later the present author (Mott 1951, last section) showed in principle how the same formula could be obtained, but with a better value of γ , by a method based on Orowan's ideas (1947). Wyatt (1953) in experimental work on pure polycrystalline aluminium and copper showed that the creep strain was the sum of a logarithmic term of type (6), with $s=1$, predominant at low temperatures and a term of the Andrade type (5) predominant at high temperatures.

In this section we shall attempt a more detailed explanation of the logarithmic term than has been given hitherto. Logarithmic creep, we believe, occurs when dislocations—e.g. those generated by Frank-Read sources or already present in the specimen—are hung up by obstacles from which thermal vibrations can release them, but when no recovery occurs of the type discussed in the next sections. If the obstacles are the islands of strain round incipient precipitates in age-hardened alloys, one expects a value of s equal to $\frac{2}{3}$; if they are other dislocations which have to be cut, then, as pointed out by Cottrell (1952), s is unity. Our approach in either case will follow that of Orowan (1947) and Mott (1951), rather than those of Mott and Nabarro (1948) and of Davis and Thompson (1950). It will ascribe the slowing down of creep to the increasing difficulty that the dislocations have in escaping from the obstacles as the material work-hardens. This theory gives values of γ much nearer to those observed.

We suppose that in a material there are, after a stress has been applied, N points per unit volume where a dislocation is held up. The activation energy to remove one is $U(\sigma)$, and the area swept out by the dislocation when it is released is A . The creep rate is then

$$d\epsilon/dt = \nu N b A \exp[-U(\sigma)/kT],$$

where ν is a frequency. If the obstacles which hold up dislocations are precipitates, Mott and Nabarro show that

$$U(\sigma) = B(1 - \sigma/\sigma_0)^{3/2}, \quad A \sim \lambda^2, \quad \nu \sim 10^{10} \text{ sec}^{-1}. \quad \dots \quad (7)$$

Here $B \sim 0.2b\lambda^2\sigma_0$ and λ is the distance between the precipitates. For dislocations held up where they have to cut other dislocations, as shown by Cottrell (1952) and in §§ 2 and 3 of this paper

$$U(\sigma) = B(1 - \sigma/\sigma_0), \quad A \sim l_0^2, \quad \nu \sim 10^{12} \text{ sec}^{-1}. \quad \dots \quad (8)$$

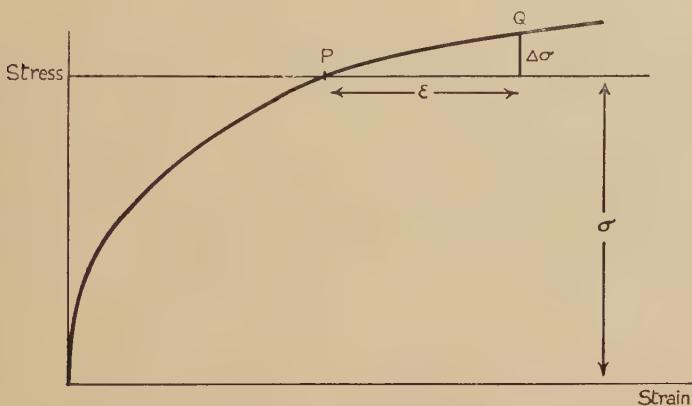
B is here the energy of a jog (c. 2 ev); and l_0 as defined in §§ 2 and 3; σ_0 is in either case the stress at which movement is possible without the help of thermal vibrations.

The exhaustion theory assumes that the values of σ_0 are spread out over a range of values, as indeed they must be, and that the dislocation traps with small σ_0 are 'exhausted' first. As was shown by Mott and Nabarro and in a simplified way by Davis and Thompson, the theory leads to a creep extension varying as $(\log \nu t)^{2/3}$ or $\log \nu t$ in the two cases (7) and (8) considered above. In the latter case, it is impossible to derive ν

from the experiments, since $\log \nu t = \log \nu + \log t$ and one cannot distinguish the constant term from the extension which occurred before creep began; Davis and Thompson on the other hand from their work on alloys could determine ν and found $c. 1 \text{ sec}^{-1}$. In view of this, following Orowan (1952), we consider that exhaustion creep, though it may occur, will mainly be included in the instantaneous extension; the phenomena observed by Wyatt and by Davis and Thompson occur because creep is slowed down due to work-hardening.

A few words may be said about the nature of this work-hardening. A movement forward of a few dislocations will in general lead to hardening even if no new ones are produced, because piled-up groups produce larger mean internal stresses than the same number dispersed at random. Therefore, although generation of new dislocations by Frank-Read sources is not ruled out in our theory, it is not necessary.

Fig. 3



Schematic work-hardening curve.

Thus in formulae (7) and (8) we take σ_0 to be the local stress required to pull a dislocation away from the obstacle that is holding it up, and neglect any variation from obstacle to obstacle. Since work-hardening is occurring, we modify formulae (7) and (8) in a way that may be understood by reference to fig. 3. At the point P the instantaneous extension terminates; here σ , the applied stress *minus* the random internal stresses just equals σ_0 , the stress required to pull the dislocations past the obstacles. As we proceed to larger strains, σ_i increases. Thus in formulae (7) and (8) we must replace $\sigma_0 - \sigma$ by the quantity $\Delta\sigma$ marked in fig. 3.

If we measure strain ϵ from the point P of fig. 3, then we may write

$$\Delta\sigma = \theta\epsilon,$$

where θ is the gradient $(d\sigma/d\epsilon)$ of the stress-strain curve at the point P. Equation (8) may thus be replaced by

$$U(\sigma) = B\theta\epsilon/\sigma_0,$$

with a corresponding equation for (7). For pure metals, then, the equation determining creep becomes

$$d\epsilon/dt = \nu N b l_0^2 \exp(-B\theta\epsilon/\sigma_0 kT). \quad \dots \quad (9)$$

Integration gives

$$\epsilon = \alpha \log(\gamma t + 1),$$

where

$$\alpha = kT\sigma_0/B\theta$$

and

$$\gamma = N b l_0^2 \nu / \alpha.$$

As regards α , the linear dependence on T is in excellent agreement with Wyatt's experimental results; the same result follows from exhaustion theories. The absolute value depends on θ , the rate of strain hardening. The observed value is $c. 10^{-3}$. As kT/B will be about 10^{-2} , this would imply that

$$\frac{\theta}{\sigma_0} = \frac{1}{\sigma} \frac{d\sigma}{d\epsilon} \sim 10,$$

which is not unreasonable. As regards the dependence on stress, this will be through θ only. Wyatt in fact finds a much smaller dependence on stress than in other forms of creep.

As regards an estimate of γ , the difficulty is to estimate N . It will certainly be much less than $1/l_0^3$, so $\gamma \ll \nu$.

Turning now to the experiments of Davis and Thompson, a similar treatment, using eqn. (7) gives

$$d\epsilon/dt = C \exp(-\beta\epsilon^{3/2}),$$

where

$$\beta = B(\theta/\sigma_0)^{3/2}/kT, \quad C = \nu N b \lambda^2.$$

In the limit of t large, integration gives approximately

$$\exp(\beta\epsilon^{3/2})/\beta^{2/3} = Ct,$$

whence

$$\epsilon = D(\log \gamma t)^{2/3}, \quad \dots \quad (10)$$

where

$$D = \beta^{-2/3} = (kT/B)^{2/3} \sigma_0 / \theta \quad \text{and} \quad \gamma = C/D.$$

Equation (10) is of the form found by Davis and Thompson. The constant B ($\sim 0.2 b \lambda^2 \sigma_0$) is larger than the equivalent quantity for the pure metal; kT/B may be $1/1000$. It will be seen that the observed values of D ($\sim 10^{-3}$) follow from quite natural values of θ/σ_0 (say 10). As regards γ ($= \nu N b \lambda^2 / D$), the difficulty again is to estimate the number N of places per unit volume where a dislocation is hung up. If we took, say, $N \sim 10^8 \text{ cm}^{-3}$ and $\lambda \sim 10^{-6} \text{ cm}$, we find $\gamma \sim 10 \text{ sec}^{-1}$. Clearly therefore quite reasonable values of N will give values of γ near to the observed one of $c. 1 \text{ sec}^{-1}$.

§ 5. THERMAL RECOVERY OF A COLD-WORKED METAL

Various surveys of the recovery process have been given; we quote from one due to Beck (1953). There are two stages in recovery; after cold-work, as already stated, there is some fragmentation of the lattice into blocks, but the blocks are strained. In the first stage the strain disappears, in the second stage the size of the blocks increases.

We have already ascribed the block formation to fine slip; the strain in the blocks, which disappears during the first stage of recovery, we ascribe to the piled-up groups of dislocations caused by coarse slip. In this section we discuss only the first stage of recovery,* which we ascribe to the dissolution of these piled-up groups by climb (Stroh 1953 a).

Various authors (Kuhlmann, Masing and Raffelsieper 1949, Cottrell and Aytekin 1950, Mott 1951) have described recovery by a differential equation of the type

$$d\sigma/dt = -A \exp \{-(W-q\sigma)/kT\}. \quad \dots \quad (11)$$

Here σ is the yield stress, and A , W , q are constants. Integration gives

$$\sigma = \sigma_0 - (kT/q) \log (t/t_0),$$

where σ_0 is the yield stress at an arbitrary time t_0 . Such an equation is not entirely satisfactory unless q is held to depend on the initial amount of cold work, because, according to Burgers (1947) using experimental results of Kornfeld (1934), the final yield stress σ that an aluminium crystal reaches on recovery at a given temperature is not independent of the initial value σ_0 , but increases with increasing σ_0 .

We shall attempt a rough model based on the model of a cold-worked metal given in paper I of this series. The flow stress of a cold-worked metal is

$$\sigma = Gbn/2\pi r,$$

where n is the number of dislocations in a piled-up group and r is the distance between such groups. We suppose that n , at any stage in recovery, decreases according to the equation

$$dn/dt = -\nu \exp (-W/kT), \quad \dots \quad (12)$$

where W is the activation energy for climb. This we take to be of the form

$$W = W_0 - \gamma \sigma b^3 n, \quad \dots \quad (13)$$

where W_0 is the activation energy for climb of a free dislocation and γ a numerical constant of order unity. It seems reasonable to suppose that a formula of this type will represent satisfactorily the reduction in the energy for climb due to the high stresses in a piled-up group.

Before recovery, according to our model, r depends on the degree of cold-work, but n does not; during recovery n decreases while r remains constant. A substance that has been heavily cold-worked and then

* Softening during the second stage is probably due to a lengthening of l , the mean length of the sources.

subjected to thermal softening is thus by no means in the same state as a substance that has been more lightly cold-worked.

Substituting for σ and n in (12), we find

$$\frac{dn}{dt} = -\nu \exp \left\{ - \left(W_0 - \frac{\gamma G b^4 n^2}{2\pi r} \right) / kT \right\}.$$

This is not quite of the form (11), containing n^2 (or σ^2) in the exponential instead of n . We can estimate the final yield stress obtained after a long anneal by setting the left-hand side equal to, say, 10^{-4} sec^{-1} ; with ν equal to 10^{12} , this gives

$$\gamma G b^4 n^2 / 2\pi r = W_0 - 40kT.$$

This gives, essentially, owing to the occurrence of n^2 on the left-hand side of this equation

$$(\sigma/\sigma_0)^2 = 1 - T/T_0, \quad 1/T_0 = 40k/W_0.$$

Here σ_0 is the yield stress after cold-work and before annealing. For aluminium, if W_0 is the activation energy for self-diffusion (37 kcal), $T_0 \sim 500^\circ$. If W_0 should include the energy of a jog in a dislocation, T_0 should be somewhat higher, but a recent paper by Stroh (1953 b) suggests that the energy of a jog in a dislocation in a piled-up group is small.

Our model predicts, in agreement with Burgers (1947) that the proportional recovery is the same for different degrees of cold work.

The model given here suggests that climb in the highly stressed region of the piled-up group is all that is necessary for recovery. As soon as a dislocation has climbed a certain small distance out of its slip plane, it can escape by the normal slip process. Subsequent polygonization of these dislocations may occur.

The roles which impurities may play in lowering the activation energy for recovery are, according to this theory, as follows :

(a) A soluble impurity, of which the atom is a different size from the matrix, can diffuse to the piled-up groups and lower the stress there. If diffusion of impurity to the piled-up groups is more rapid than the 'climb' of the dislocations, then such a process will clearly raise the activation energy for climb.

(b) Insoluble impurities (e.g. oxides), if they lower the slip distance, will decrease n , and so the activation energy.

On the other hand self-diffusion may be easier along the boundaries between two materials. Also, if the slip lines normally terminate on grain boundaries, this should lower the activation energy for recovery, because diffusion is quicker along the grain boundaries.

§ 6. STEADY-STATE CREEP

The explanation of creep at a constant velocity is likely to be similar to that for recovery; the creep rate should be determined by equating the rates of work-hardening and of recovery (see, for instance, Cottrell

and Aytekin 1950). On such a model one takes a formula such as (11) for the rate of recovery and equates it to the rate of hardening $\theta d\epsilon/dt$, where θ is the gradient of the stress-strain curve already defined. Thus

$$\frac{d\epsilon}{dt} = \frac{A}{\theta} \exp \left\{ - \frac{(W - q\sigma)}{kT} \right\}, \quad \dots \quad (14)$$

an equation of the type proposed by several investigators (Kauzmann 1941, Nowick and Machlin 1947 and Feltham 1952). In the interpretation of this formula, however, we cannot use quite the recovery mechanism discussed in § 5. If the considerations of § 2 are correct, slip in creep is not dynamic, and we cannot expect to have anything like 1000 dislocations piled up against one barrier; probably at any one time there may not be more than one or two. Therefore, we cannot identify q with nb^3 , for, if we did, values of n from 500 to 1000 would be needed to explain the values of q observed for instance by Feltham. We need another mechanism of stress magnification, which will lower the activation energy of climb.

For this we suggest the following: suppose that the barriers are sessile dislocations. These dislocation lines will not, in general, lie for a distance L all in one plane; we may expect them to extend over a smaller distance l , say 10^{-4} cm or less, and then move into another plane or branch into two planes. The barriers against which the dislocations pile up will thus have gaps in them, of length l . At the ends of these gaps we may expect a stress magnification of order l/b for a dislocation held up by a barrier. It is here that the 'climb' begins which enables the dislocations to escape from the barriers. We thus identify q with lb^2 , which gives a term of the right order.

Our picture of creep is thus as follows: A relatively small number of sources generates dislocations by the non-dynamic process, and these are held up by barriers, so producing small groups which increase the yield stress as described in § 2. Owing to gaps in the barriers the dislocations can 'climb' away from them, this being the rate determining process. Thereafter they polygonize, producing the cells observed by Wood and his co-workers (cf. Wood and Suiter 1952, Rachinger 1952) and by McLean (1952, 1953). McLean finds that for large creep strains the angle between polygonized elements is equal to the creep strain. This suggests

(a) that the size of the polygonized blocks is determined by and about equal to the slip distance L , and

(b) that, when the strain is large, only half the polygonized elements are slipping at all fast, namely those in which the stress is largest.

If this is so, a dislocation having escaped from its barrier will move to the nearby boundary, and each boundary will receive dislocations of one sign only, which explains the above result.

We now have to ask what is the nature of the boundaries. It is unlikely that they are those existing before deformation, and indeed

the block size depends on temperature and strain rate. We suggest that they are formed because, in a polycrystalline cubic metal, there will always be slip on a number of non-parallel planes. This, as we know, produces sessile dislocations.

It is possible, also, that the grain boundaries may act as barriers if the grain diameter is smaller than L . Under these conditions we should expect :

(a) A value of W in formula (14) equal to the activation energy for grain boundary diffusion, instead of the value for volume diffusion, which is what we expect for climb within the grain. This makes for more rapid creep.

(b) A much smaller dependence of creep rate on stress ; we cannot envisage any stress-raising mechanism at the grain boundary. The dependence on stress should be as $\sinh(nb^3\sigma/kT)$, which for fine slip will be as $n\sigma$ and hence as σ^2 .

The model given here suggests that the ideal creep-resistant material would be of one of fine grain-size in which the grain boundaries were filled with oxide or carbide or some substance adhering sufficiently strongly to the grain for the grain boundary diffusion to be small. Perhaps the oxide-rich sintered aluminium powders described by Irmann (1952) belong to this category.

We may remark here on the interesting results of Rachinger (1952), who reports that after deformation of polycrystalline aluminium by 50%, for instance at 300°C at 0.1% per hour, the grains remained equiaxed. We suggest that that slip here is by fine slip ; dislocations are moved into the grain boundaries along a high proportion of atomic planes, and the acceptance of so many dislocations enables the grain boundaries to migrate during creep, thus retaining their original form, which for an annealed material will represent a minimum of the free energy. We do not think that grain boundary slip will account for more than a small proportion of these large strains.

§ 7. ANDRADE CREEP

We have already described a form of transient creep which occurs, we believe, when there is no recovery in the sense described in § 5. This theory gave a logarithmic dependence of strain on time. Andrade (1910) was the first to propose a variation of ϵ with $t^{1/3}$, and to show that for a wide variety of polycrystalline metals the transient creep obeys this law. We believe that this type of creep, like that described in the last section, is associated with recovery, or 'climb' of dislocations, and, like logarithmic creep, it is characteristic of that part of the stress-strain curve where the flow stress slightly exceeds the applied stress (fig. 3). As Andrade (1952) has shown, after prior rapid strain there is no transient creep.

Referring then to fig. 3, we suppose that at the point P, at the end of the instantaneous extension, the applied stress is equal* to the internal stress σ_i . A steady state is only reached when σ slightly exceeds σ_i and transient creep occurs till this state is reached. The reason is as follows: each dislocation which diffuses away from any piled-up group and, perhaps after some movement by slip, eventually joins a polygon boundary, will change the stress in the neighbourhood. Let us then consider the stress at a given point X in the crystal. Each time that a dislocation escapes from a nearby piled-up group, let the stress there change by, say, σ_1 . Let these changes be of random sign. Then after n such events, the stress will have changed by $n^{1/2}\sigma_1$. Now suppose that at the point X there is a Frank-Read source, and that the state of the crystal is represented by the point Q on the stress-strain curve in fig. 3. Then the number n of such events required to raise the stress at X by $\Delta\sigma$, after which the source will produce another dislocation, is $(\Delta\sigma/\sigma_1)^2$. This as before may be written $(\theta\epsilon/\sigma_1)^2$. Thus if ω is the frequency with which dislocations escape from nearby sources, the creep rate is given by

$$d\epsilon/dt = NL^2b\omega(\sigma_1/\theta\epsilon)^2,$$

where N is the number of sources per unit volume, and L the slip distance. Integration gives

$$\epsilon = \beta t^{1/3},$$

where

$$\beta^3 = 3NL^2b\omega(\sigma_1/\theta\epsilon)^2.$$

This is of the required form. The derivation could be adapted to any mechanism in which work-hardening and recovery produce local fluctuations of stress.

It will be reasonable to write

$$\sigma_1 \sim Gb/l,$$

where l is of the order of the distance between dislocations, say 10^{-4} cm. The frequency ω is supposed to be due to the same process as determines steady state creep, namely the escape of dislocations from piled-up groups. We may thus set

$$\omega = v \exp \{-(W - q\sigma)/kT\}$$

where W and q have the same values as for steady-state creep. Following an earlier estimate we set $\theta \sim 10Gb/l$. We then find

$$\beta^3 \sim 0.01NL^2b\nu \exp \{-(W - q\sigma)/kT\}.$$

Thus we may write

$$\beta^3 \sim 0.01\dot{\epsilon}_s$$

where $\dot{\epsilon}_s$ is the creep rate when this has become constant. Another way of expressing the equation of transient creep is

$$d\epsilon/dt \sim \dot{\epsilon}_s/100\epsilon^2,$$

* Neglecting the stress Gb/l required to operate the sources.

which suggests that the transient flow should approach the steady rate when $\epsilon \sim 0.1$. This estimate is, of course, very rough.

Some preliminary results of Feltham's (1953) suggest that for some of the carbon steels investigated in the γ -phase, the prediction that β^3 varies with σ and T as does $\dot{\epsilon}_s$ is roughly verified.

Of course many assumptions are made in the above derivation. One is that the density of barriers is independent of creep rate. This may not be the case, especially since slip on more than one plane is occurring.

§ 8. VACANCIES CREATED BY DEFORMATION

Seitz (1952) has reviewed the evidence that moving dislocations form vacancies, and has proposed several mechanisms by which this might occur. One of these has been discussed in detail by the present author (1952) and some of its consequences pointed out. Here we shall make some further points about this hypothesis :

(a) Seitz deduces from the change of electrical resistance of a cold-worked metal that the number of vacancies per atom is about $10^{-4}\epsilon$, where ϵ is the strain. The mechanism discussed by the present author gives, in fact, $\epsilon b/l$. It depends however, on the hypothesis that the motion is dynamic. For 'fine slip', if the temperature is high enough for the vacancies to disperse as soon as they are formed, we should expect on the same model $\epsilon bL/l^2$, or 100 times as many.

(b) In the author's paper (1952), it was suggested that the vacancies formed in the neighbourhood of a slip band produce some kind of local annealing which accounts for the observed 'clustering' of the elementary slip lines and also for the drop in the strain hardening curve as the temperature is raised. It was suggested that this softening consists in the dissolution of whatever piled-up groups are nearest to the slip plane, not particularly those at the end of it ; and that some of the secondary groups already referred to might be those primarily affected. A table was given (Mott 1952, p. 1177) in which the activation energy for the motion of vacancies was estimated as 16–18 kcal for Cu, Ag and Au, values in rather good agreement with those obtained by Manintveld (1952) from resistivity measurements.

There is other evidence, apart from clustering of slip lines, that something of the sort occurs. In the author's previous paper, the behaviour of a metal strained at liquid air temperatures and then at room temperature was quoted (for experimental results see Dorn, Goldberg and Tietz 1949). The speed with which the metal hardens after the drop in temperature, and the fact that, if the initial strain is small, the curve for straining at low temperatures is quickly reached, were taken to show that a very localized region round each slip plane had been softened during deformation at room temperature. The further evidence referred to is provided by the work of Cherian, Pietrokowsky and Dorn (1949) on the annealing of cold-worked aluminium. They find that after a 9% strain and anneal at 32°C or 100°C, the material on further straining returns quickly to the

original stress-strain curve. This seems to us a phenomena of the same type ; the moving dislocations leave behind them, after the original deformation, some type of debris* (interstitial atoms?), which is immobile at room temperature but can move between 30°C and 100°C ; and this allows local recovery.

According to these authors, however, after an anneal at 150°C or 205°C, the original stress-strain curve is not regained. At these temperatures, we suppose, vacancies are produced thermally, and these allow some dissolution of the groups of piled up dislocations and with it a diminution of σ_i throughout the material. The authors quoted find an activation energy (33 kcal) for this type of recovery very close to the activation (37 kcal) for self-diffusion.

(c) In a recent paper Buffington and Cohen (1952) have shown that, if iron at 890°C is deformed at a rate of 25% per hour, there is a large increase in the coefficient of self-diffusion. This they ascribe to a large increase in the number of vacancies. Any effect of this kind will obviously be of importance for technical materials where precipitation is taking place during creep. Also we need to consider whether these vacancies themselves will influence 'climb' and thus the creep rate. We shall come to the conclusion that they do not.

The authors quoted found that the self-diffusion coefficient during creep could be expressed in the form

$$D_s = D_0(1 + 180,000\dot{\epsilon}).$$

Let us suppose, following Seitz, that the rate per unit volume of creation of vacancies is $f\dot{\epsilon}/a^3$, where a is the inter-atomic distance. Seitz's figure for f is 10^{-4} ; we suggest above that for fine slip it may be 100 times bigger. According to Seitz's mechanism interstitial atoms are also formed, and the two will recombine ; in the absence of any more effective way of getting rid of vacancies, the number n/a^3 per unit volume will be given by

$$f\dot{\epsilon} = \nu n^2 \exp(-U/kT) \dots \dots \dots \quad (15)$$

which gives n proportional to $\sqrt{\dot{\epsilon}}$ instead of $\dot{\epsilon}$. We must thus assume the presence of sinks for vacancies with concentration higher than that of the interstitial atoms. Formula (15) then gives an upper limit to the number of interstitials and thus to the number of vacancies. In the experiments quoted the maximum creep rate was 10^{-4} sec^{-1} . Putting $f=10^{-2}$, $\nu=10^{12} \text{ sec}^{-1}$, this gives $n=10^{-9} \exp(\frac{1}{2}U/kT)$. The number of

* Marx, Cooper and Henderson (1952) find a whole range of activation energies from 0.2 ev upwards for the motion of this debris in copper, as shown by the recovery of the electrical resistance after cold-work. Manintveld (1952) also finds, as well as the value of about 0.8 ev for Cu quoted above, partial recovery with an activation energy of 0.2 ev. This may be due to pairs of vacancies. Bowen, Eggleston and Kropschot (1952) find for the recovery of the resistance of copper an activation energy of 28.3 kcal. This may be due to the motion of interstitial atoms that we postulate here to account for the result of Cherian *et al.*

vacancies present in thermal equilibrium is $\exp(-W/kT)$. At the above creep rate the self-diffusion coefficient was five times the normal value. Thus the experiments can be explained along these lines only if

$$10^{-9} \exp(\frac{1}{2}U/kT) \gg 5 \exp(-W/kT), \dots \quad (16)$$

or

$$\exp[-(W + \frac{1}{2}U)/kT] \ll 10^{-10}.$$

But the coefficient of self-diffusion itself at this temperature is $c. 6 \times 10^{-12} \text{ cm}^2 \text{ sec}^{-1}$; and if we write this in the form $\nu a^2 \exp[-(W+U)/kT]$ and take $\nu a^2 \sim 10^{-3}$, we see that $\exp[-(W+U)/kT] \sim 10^{-9}$. Thus the inequality (16) is quite impossible for any positive U . We cannot therefore account for the required number of vacancies by the Seitz mechanism.

We shall now suggest an additional mechanism, by which holes but not interstitials are produced during recovery or creep. At high temperatures a dislocation is normally absorbing and giving out holes from its jogs; when, under the influence of a stress gradient, it 'climbs', it gives out more than it receives or vice versa. If the stress is very strong, however, it is clear that it can move much faster in one direction than the other; for there is no limit to the rate at which it gives out vacancies, but for movement in the other direction its speed is limited by the number of vacancies present. We suggest, therefore, that under the high stresses existing in the piled-up groups, dislocations normally climb in such a direction as to create vacancies. The occurrence in formulae for the creep rate of terms of the type $\sinh(q\sigma/kT)$, where $q\sigma/kT$ is larger than unity, suggests that this is the case.

The rate of production of vacancies by this mechanism will clearly be proportional to the creep rate. Moreover, the vacancies can only disappear by combining with dislocations or grain boundaries. Thus a linear relation between n and $\dot{\epsilon}$ is to be expected.

It is possible that this production of vacancies during recovery is one cause of tertiary creep in age-hardened alloys, the vacancies accelerating diffusion and so over-ageing. If this is so, we should expect tertiary creep to occur, for varying stress, at roughly constant strain. There are of course other causes of tertiary creep, such as recrystallization (Greenwood and Worner 1939, Hirst 1940).

§ 9. A REVIEW OF THE DYNAMIC HYPOTHESIS

The theories put forward in this paper suggest a reconsideration of the dynamic model of slip on which paper I was based. There it was stated that if the movement of dislocations is undamped or subject to insufficient damping to prevent them approaching the speed of sound, a source once started will go on generating dislocations until the stress drops to zero. This is not quite correct; the more exact estimate of Fisher, Hart and Pry (1952) suggests that the stress can drop to about

one-half of its original value (Gb/l) before generation by the source comes to an end. None of the qualitative conclusions of paper I are affected by this.

Fisher, Hart and Pry consider that the action of a source may be stopped by the stress from the moving dislocations which it generates. It has been pointed out to the author by Dr. A. Seeger (private communication) that there is a mistake in their paper, and that the assumptions that they make give a value of 6 rather than 300 for n , the number of dislocations that can be formed. On the other hand, since their theory gives $\log n$ rather than n and little is known about the dynamic movement of dislocations, little reliance can be placed on their numerical estimate. The true value of n , however, is likely to be less than 1000.

On the other hand, if there is even weak Cottrell locking of the source, the stress required to operate it will be greater and perhaps much greater than Gb/l . Under these conditions it is probable that the moving dislocations will not produce a back stress big enough to stop the source. Moreover we think it probable that some kind of Cottrell locking* is, in aluminium at any rate, normally responsible for the yield of single crystals. This is shown, as pointed out in paper I, by the large temperature dependence of the yield point observed by Rosi and Mathewson (1950), which can hardly be explained on any other hypothesis.

A further result of Cottrell locking is that there should be little or no fine slip. This is probably the reason why Chen and Mathewson (1951) do not observe deformation bands in α -brass, where Cottrell locking probably occurs. We consider that deformation bands, at least in the early stages, are formed by fine slip.

We turn now to the consideration of the results of Wilsdorf and Kuhlmann-Wilsdorf (1951 and 1952), who have examined, using a silica replica, the slip lines on an electropolished surface of aluminium. They find much evidence of fine slip and have shown that the coarse slip lines can be resolved into a cluster of lines near together, each of height $c. 300 \text{ \AA}$. As already stated, we do not consider that the electropolished surface gives reliable results about slip in the interior; following Hollomon we suppose that dislocation lines which end on the surface will act like sources of twice their actual length. These will then generate dislocations by the non-dynamic mechanism. It follows that a layer of thickness L (say 10^{-1} to 10^{-2} cm) will be subjected only to slow, linear hardening. Thus no dynamic or coarse slip can be initiated there. But if a source just outside this region starts to generate dislocations dynamically, and a coarse slip band ends at a barrier within the soft region, at E in fig. 4,

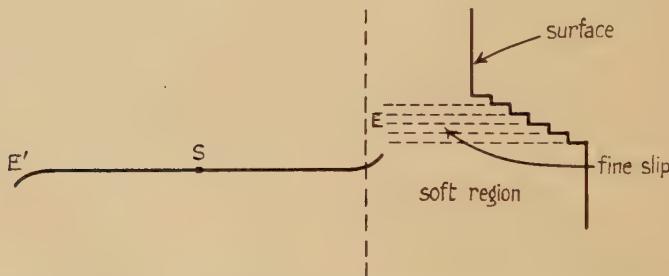
* Cottrell locking may be understood in two senses: the diffusion of impurities to the dislocation lines of the Frank-Read sources, which will give strong locking, and the weak locking by impurities distributed at random in the lattice which may accidentally find themselves in the path of the dislocations, as described by Mott (1952 b) and Friedel (1953).

it is likely enough that the stresses round E will be relieved by fine slip in the manner shown. In this way we suggest that the observations of Wißdorf *et al.* may be explained.

§ 10. A POSSIBLE EXPLANATION OF COARSE AND FINE SLIP IF THE DAMPING DUE TO SOUND WAVES IS TOO GREAT FOR DYNAMIC MOVEMENT TO OCCUR

As already stated, the work of Nabarro (1951) suggests that the motion of dislocations may be damped by sound waves, and it is not yet certain that, for the stresses at which metals yield, dislocations can approach the speed of sound. If they cannot, the explanation of coarse slip given by Mott (1952 a) and Fisher, Hart and Pry must be modified. The general conclusions of this paper can be shown, however, to follow from a slightly different hypothesis. This is that the yield point, or stress at which a source acts, is normally determined by impurities. It is not suggested that the impurities necessarily diffuse towards the dislocation; but that

Fig. 4



Suggested mechanism of slip band formation on an electro-polished surface. S is the source, EE' a slip line in the interior.

the dislocation line, in its state of lowest energy, is slightly bent, so as to relieve the stress round as many impurities as possible. An analysis of this type of locking has been given by Mott (1952 b, 1953), which suggests that the yield stress due to impurities present in concentration c will be of order, say, $0.1 Gc^{2/3}$, so that a concentration of 30 parts in a million will give a locking stress greater than Gb/l when $l \sim 10^{-4}$ cm.

Now a dislocation moving with quite a small fraction of the speed of sound (say one-tenth) will not be subject to this kind of locking; so once a dislocation has broken away, it will, as in paper I, continue generating dislocations until the applied stress drops to Gb/l . The qualitative results of paper I follow. But if the dislocation is temporarily held up by barriers such as crossing dislocations, as in § 2 of this paper, it will have time to settle down into its equilibrium form and become locked again, so that the results of this paper follow too.

If this explanation of coarse slip is correct, it would follow that very pure metals should not show coarse slip and that hardening should be linear in the strain, even in the polycrystalline state.

APPENDIX

In this section we shall investigate the behaviour of a dislocation, moving with velocity V , which impinges on an obstacle. We have to ask, for what energy will it break through the obstacle?

We suppose that the dislocation line is parallel to the x -axis and that it is moving with velocity V parallel to the y -axis. For simplicity we suppose that it impinges on two obstacles at the points $(\pm \frac{1}{2}l, 0)$. Then its equation of motion after the impact will be

$$\rho \frac{\partial^2 y}{\partial t^2} = T \frac{\partial^2 y}{\partial x^2},$$

where $T = \frac{1}{2}Gb^2$, $\rho \sim M/b$ and M is the mass of an atom. After impact we set

$$y = \sum_n A_n \cos \{(2n+1)\pi x/l\} \sin \{(2n+1)\pi c t l\},$$

where $c^2 = T/\rho$.

The initial condition we take to be that dy/dt is equal to V everywhere; this gives

$$A_n = (-1)^n V l / c \pi^2 (n + \frac{1}{2})^2.$$

A short calculation shows that if $\zeta (= \frac{1}{2}l - x)$ is the distance from one end, the form of the dislocation at time t near the end is given by the equation

$$y = \frac{2V\zeta}{\pi c} \sum_n \frac{\sin \{(n + \frac{1}{2})(2\pi c t / l)\}}{n + \frac{1}{2}}.$$

On summing, this turns out to be independent of t , and gives

$$y = (V/c)\zeta.$$

Thus, immediately after impact, the dislocation makes an angle θ of tangent V/c with the x -axis, and at the obstacle this does not change with time.

Immediately after impact, then, the force on the obstacle is $2VT/c$. If as before αGb^3 is the energy of a jog, the dislocations should break through if

$$2VTb/c > \alpha Gb^3.$$

Putting $T = \frac{1}{2}Gb^2$, this gives

$$V/c > \alpha.$$

V will certainly be of the order of c , but according to Stroh's estimate (§ 3), for close-packed structures α is greater than unity. It looks therefore as if, for these structures at any rate, the dislocation should not break away. The whole calculation, however, is made without consideration of the 'relativistic' corrections which arise when a dislocation approaches the speed of sound, and must be regarded as tentative.

REFERENCES

ANDRADE, E. N. DA C., 1910, *Proc. Roy. Soc. A*, **84**, 11 ; 1914, *Ibid.*, **90**, 329 ; 1952, *J. Iron Steel Inst.*, **171**, 217.

ANDRADE, E. N. DA C., and HENDERSON, C., 1951, *Trans. Roy. Soc. A*, **244**, 177.

ANDRADE, E. N. DA C., and ROSCOE, R., 1937, *Proc. Phys. Soc.*, **49**, 152.

BECK, P., 1953, *Acta Metallurgica* (in press).

BLEWITT, T. H., 1953, *Acta Metallurgica* (in press).

BOWEN, D., EGGLESTON, R. R., and KROPSCHOT, R. H., 1952, *J. Appl. Phys.*, **23**, 630.

BROWN, A. F., 1951, *J. Inst. Met.*, **80**, 115 ; 1952, *Advances in Physics*, **1**, 427.

BROWN, A. F., and HONEYCOMBE, R. W. K., 1951, *Phil. Mag.*, **42**, 1146.

BUFFINGTON, F. S., and COHEN, M., 1952, *Trans. A.I.M.E.*, **194** (in *J. Metals*, **4**), 1085.

BURGERS, W. G., 1947, *Proc. Acad. Sci. Amst.*, **50**, 452.

CHERIAN, T. V., PIETROKOWSKY, P., and DORN, J. E., 1949, *Trans. A.I.M.E.*, **185** (in *J. Metals*, **1**), 948.

CHEN, N. K., and MATHEWSON, C. H., 1951, *Trans. A.I.M.E.*, **191** (in *J. Metals*, **3**), 653.

CHEN, N. K., and POND, R. B., 1952, *Trans. A.I.M.E.*, **194** (in *J. Metals*, **4**), 1085.

COTTRELL, A. H., 1952, *J. Mech. Phys. Solids*, **1**, 53.

COTTRELL, A. H., and AYTEKIN, V., 1950, *J. Inst. Met.*, **77**, 389.

DAVIS, M., and THOMPSON, N., 1950, *Proc. Phys. Soc. B*, **63**, 847.

DORN, J. E., GOLDBERG, A., and TIETZ, T. E., 1948, *Met. Tech.*, **15** (b), and 1949, *Trans. A.I.M.E.*, **180**, 205.

FELTHAM, P., 1952, *Nature, Lond.*, **169**, 976 ; 1953, *Proc. Phys. Soc. B*, **66**, (in the press).

FISHER, J. C., HART, E. W., and PRY, R. H., 1952, *Phys. Rev.*, **87**, 958.

FRANK, F. C., 1950, *Symposium on Plastic Deformation of Crystalline Solids* (Pittsburgh : Carnegie Inst. of Tech. and Office of Naval Research).

FRENCH, R. S., and HIBBARD, W. R., 1950, *Trans. A.I.M.E.*, **188** (in *J. Metals*, **2**), 53.

FRIEDEL, J., 1953, *Proc. Phys. Soc.* (in press).

GAY, P., and KELLY, A., 1953, *Acta. Crystallogr.*, **6**, 165.

GOUGH, H. J., and WOOD, W. A., 1936, *Proc. Roy. Soc. A*, **154**, 510 ; 1938, *Ibid.*, **165**, 358.

GREENWOOD, J. N., and WORNER, H. K., 1939, *J. Inst. Met.*, **64**, 135.

HANSON, D., and WHEELER, M. A., 1931, *J. Inst. Met.*, **45**, 229.

HEDGES, J. M., and MITCHELL, J. W., 1953, *Phil. Mag.*, **44**, 223.

HEIDENREICH, R. D., 1951, *Bell Syst. Tech. J.*, **30**, 867.

HEIDENREICH, R. D., and SHOCKLEY, W., 1948, *Bristol Conference on the Strength of Solids* (London : Physical Society), p. 57.

HERRING, C., 1950, *J. Appl. Phys.*, **21**, 437.

HIRSCH, P. B., 1952, *Acta. Crystallogr.*, **5**, 172.

HIRST, H., 1940, *Proc. Australian Inst. Min. Metallurgy*, **118**, 101.

HOLLOWOM, J. H., 1945, *Trans. Am. Inst. Min. Met. Eng.*, **162**, 268 ; 1952, *Report to Solvay Conference*.

HONEYCOMBE, R. W. K., 1950, *Proc. Phys. Soc. A*, **63**, 672.

IRMANN, R., 1952, *Metallurgia*, **6**, 125.

JAOU, B., and CRUSSARD, C., 1952, *C. R. Acad. Sci., Paris*, **234**, 700.

KAUZMANN, W., 1941, *Trans. Amer. Inst. Min. Met. Eng.*, **143**, 57.

KORNFIELD, M., 1934, *Phys. Z. Sowjet.*, **6**, 329.

KUHLMANN, D., MASING, G., and RAFFELSPIEGER, J., 1949, *Z. Metallkunde*, **40**, 241.

KUPLMANN-WILSDORF, D., VAN DER MERWE, J. H., and WILSDORF, H., 1952, *Phil. Mag.*, **43**, 632.

LEIBFRIED, G., 1950, *Z. Phys.*, **127**, 580.
LUCKE, K., and LANGE, H., 1952, *Z. Metallkunde*, **43**, 55.
MANINTVELD, J. A., 1952, *Nature, Lond.*, **169**, 623.
MARX, J. W., COOPER, H. G., and HENDERSON, J. W., 1952, *Phys. Rev.*, **88**, 106.
MCLEAN, D., 1952, *J. Inst. Met.*, **80**, 507; 1953, *Ibid.*, **81**, 287.
MOTT, N. F., 1951, *Proc. Phys. Soc. B*, **64**, 729; 1952 a, *Phil. Mag.*, **43**, 1151; 1952 b, *Imperfections in Nearly Perfect Crystals*, Shockley ed. (New York : Wiley), p. 173; 1953, *Proc. Roy. Soc. A* (in the press); Bakerian Lecture.
MOTT, N. F., and NABARRO, F. R. N., 1948, *Bristol Conference on the Strength of Solids* (London : Physical Society), p. 1.
NABARRO, F. R. N., 1948, *Bristol Conference on the Strength of Solids* (London : Physical Society); 1951, *Proc. Roy. Soc. A*, **209**, 278.
NISHIMURA, H., and TAKAMURA, J., 1952, *Tech. Rep. Eng. Res. Inst., Kyoto Univ.*, **2**, 139.
NOWICK, A. S., and MACHLIN, E. S., 1947, *J. Appl. Phys.*, **18**, 79.
OROWAN, E., 1947, *J. West Scotland Iron Steel Inst.*, **54**, 45; 1952, *Imperfections in Nearly Perfect Crystals*, Shockley ed. (New York : Wiley), p. 191.
RACHINGER, W. A., 1952 a, *Bull. Inst. Met.*, **1**, 125 and *J. Inst. Met.*, **80**, 415; 1952 b, *J. Inst. Met.*, **81**, 33.
ROSI, F. D., and MATHEWSON, C. H., 1950, *Trans. A.I.M.E.*, **188** (in *J. Metals*, **2**), 1159.
SEITZ, F., 1952, *Advances in Physics*, **1**, 43.
SMITH, G. C., and DEWHIRST, D. W., 1949, *Research*, **2**, 492.
STROH, A. N., 1953 a, *Proc. Roy. Soc. A* (in the press); 1953 b, *Proc. Phys. Soc.* (in the press).
UDIN, H., SHALER, A. J., and WULFF, 1949, *Trans. A.I.M.E.*, **185** (in *J. Metals*, **1**), 186.
WARREN, B. E., and AVERBACH, B. L., 1952, *J. Appl. Phys.*, **23**, 497.
WILSDORF, H., and KUHLMANN-WILSDORF, D., 1951, *Naturwiss.*, **38**, 502; 1952, *Z. f. angewandte Physik*, **10**, 361.
WOOD, W. A., and SUITER, J. W., 1952, *J. Inst. Met.*, **80**, 501.
WYATT, O., 1953, *Proc. Phys. Soc. B*, **66** (in press).

LXXIX. *Pairing Energy in the $j-j$ Coupling Model*

By A. J. M. HITCHCOCK
Cavendish Laboratory, Cambridge*

[Received March 2, 1953]

ABSTRACT

The pairing energy of two nucleons of the same kind in a shell has been calculated. Harmonic oscillator wave-functions and very short-range nuclear forces were assumed. It is found that the empirical predictions of Mayer (1950) are verified in all cases but one and that the exception accounts for the anomalous spin of ^{137}Ba . Certain new peculiarities are predicted (e.g. in quadrupole moments and the spin of ^{19}F) which are found to be in satisfactory agreement with experiment. The averaged variation of pairing energy with A is found not to agree with Bohr and Wheeler's semi-empirical formula.

§ 1. DEFINITION OF PAIRING ENERGY

M. G. MAYER (1950) has shown that, assuming a δ -function interaction between nucleons, the energy of m nucleons of the same kind in the shell $(nl_j)\dagger$ is given by

$$\begin{aligned} E = & -m(2j+1)I_{nl}, \text{ } m \text{ even,} \\ & -(m-1)(2j+1)I_{nl}, \text{ } m \text{ odd,} \end{aligned} \quad \dots \dots \dots \quad (1)$$

where I_{nl} is proportional to

$$\int_0^\infty |R_{nl}|^2 r^2 dr$$

and R_{nl} is the radial wave-function.

Thus, neglecting factors which are independent of n , l , and j , as we shall throughout this work, we may define the pairing energy, P , as

$$P = (2j+1)I_{nl}. \quad \dots \dots \dots \quad (2)$$

With forces of extended range it is no longer possible to regard the pairing energy purely as a function of the shell because it now depends on the number of nucleons in that shell. It is reasonable to suppose that (2) continues to give a useful approximation.

The form of R_{nl} is more important (in that it has a greater influence on the result) than the radial dependence of the nuclear forces. Here, harmonic oscillator functions will be used: since these are good approximations, we assume that the error introduced by their use will be small.

* Communicated by the Author.

† Following Mayer, n is taken to be (no. of nodes in radial wave-function + 1).

This may not be so, however, since the use of hydrogen-like wave-functions induces differences by factors of 10^2 , but this is a very extreme case.

Returning now to the form (2), we take

$$R_{nl} = r^l \exp(-\beta r^2) H_{nl}(\beta^{1/2}r),$$

where H_{nl} is the appropriate polynomial and β depends on the nuclear radius.

Then

$$\begin{aligned} I_{nl} &= \frac{\int_0^\infty |R_{nl}(r)|^4 r^2 dr}{\left(\int_0^\infty |R_{nl}(r)|^2 r^2 dr\right)^2} = \beta^{3/2} \frac{\int_0^\infty (\exp(-x^2) x^l H_{nl}(x))^4 x^2 dx}{\left(\int_0^\infty (\exp(-x^2) x^l H_{nl}(x))^2 x^2 dx\right)^2} \\ &= \beta^{3/2} J_{nl}, \text{ say.} \end{aligned}$$

$$\text{Also, } A^{2/3} = (\text{av. value of } r^2) = \frac{\int_0^\infty |R_{nl}(r)|^2 r^4 dr}{\int_0^\infty |R_{nl}(r)|^2 r^2 dr} = \frac{\nu + \frac{3}{2}}{\beta}$$

where $\nu = (l + 2n - 2)$ is the usual radial quantum number for harmonic oscillators.

Hence

$$\beta^{3/2} = (2\nu + 3)^{3/2}/A$$

and

$$AP = (2j+1)(2\nu+3)^{3/2} J_{nl} \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad (3)$$

(where, as always, constant factors are omitted).

By simple algebra :

$$J_{1l} = 2^{-2l} (4l+1)!! [(2l+1)!!]^{-2}, \text{ where } (2a+1)!! = 1 \cdot 3 \cdot 5 \dots (2a+1) \quad (4)$$

$$J_{2l} = 2^{-2l-6} (4l+1)!! [(2l+3)!!]^{-2} (64l^2 + 176l + 123).$$

These values are tabulated in table 1.

§ 2. ANOMALOUS CASES

The table shows that in regions where two levels compete, so that the variation with A is irrelevant, the pairing energy is greater in the level of greater j , except in the region $76 < (N \text{ or } Z) < 82$, where the $2d_{3/2}$ and the $3s_{1/2}$ levels compete. The experimental position is shown in table 2. The $3s_{1/2}$ neutrons fill in pairs, but the $3s_{1/2}$ protons do not. The difference is presumably due to the smaller pairing energy in the larger nuclei.

Table 1 also shows that the pairing energy difference ΔP , between the competing shells $1f_{5/2}$ and $2p_{3/2}$ is predicted to be very small. The signs of the quadrupole moments* (table 3) indicate that the configurations for 29, 31, 33 and 35 protons are not the usual $(p_{3/2})$, $(f_{5/2})^2 (p_{3/2})$, $(f_{5/2})^4 (p_{3/2})$, $(f_{5/2})^6 (p_{3/2})$, but $(p_{3/2})$, $(p_{3/2})^3$, $(f_{5/2})^2 (p_{3/2})^3$, $(f_{5/2})^4 (p_{3/2})^3$; which means that $\Delta S < \Delta P < 2\Delta S$, ΔS being the shell energy difference.

* The argument from the sign of Q is due to Jensen. (Private communication from M. G. Mayer.)

Table 1. Pairing Energies of Shells

Brackets { } indicate levels in competition.
 Brackets () indicate levels filled in pairs only.

N or Z	Shell	ν	J_{nl}	$AP = (2j+1)(2\nu+3)^{3/2} J_{nl}$
1-2	$1s_{1/2}$	0	1.000	10.4
3-6	$1p_{3/2}$	1	0.416	18.6
7-8	$1p_{1/2}$	1	0.416	9.3
8-16	{ $1d_{5/2}$ $2s_{1/2}$	2	0.266 } 2 0.640 }	29.6 } 23.7 }
17-20	$1d_{3/2}$	2	0.266	19.7
20-28	$1f_{7/2}$	3	0.194	41.9
29-38	{ $1f_{5/2}$ $2p_{3/2}$	3	0.194 } 3 0.283 }	31.4 } 30.6 }
39-50	{ $2p_{1/2}$ $1g_{9/2}$	3	0.283 } 4 0.156 }	15.3 } 56.9 }
51-76	{ $1g_{7/2}$ $2d_{5/2}$ ($1h_{11/2}$)	4	0.156 } 4 0.184 } 5 0.126 }	45.5 } 40.3 } 70.9 }
77-82	{ $2d_{3/2}$ $3s_{1/2}$	4	0.184 } 4 0.504 }	26.8 } 36.8 }
83-126	{ $1h_{9/2}$ $2f_{7/2}$ $2f_{5/2}$ $3p_{3/2}$ $3p_{1/2}$ ($1i_{13/2}$)	5	0.126 } 5 0.139 } 5 0.139 } 5 0.228 } 5 0.228 } 6 0.107 }	59.1 } 52.1 } 39.1 } 42.7 } 21.4 } 87.0 }

Table 2. Shell of Odd Nucleon in Region $76 < (N \text{ or } Z) < 82$
 Deduced from spins and magnetic moments tabulated by Mack (1950).

Nucleus	N	Z	Shell
^{191}Ir	114	77	$2d_{3/2}$
^{193}Ir	116	77	$2d_{3/2}$
^{197}Au	118	79	$2d_{3/2}$
^{203}Tl	122	81	$3s_{1/2}$
^{205}Tl	124	81	$3s_{1/2}$
^{131}Xe	77	54	$2d_{3/2}$
^{135}Ba	79	56	$2d_{3/2}$
^{137}Ba	81	56	$2d_{3/2}$

Thus ΔP is rather smaller than most pairing energy differences in agreement with the theoretical prediction, although it is not as small as is predicted to be.

Table 3. Spins and Signs of Quadrupole Moments, $28 < Z < 38$

Data from Mack (1950).

Nucleus	Z	Spin	Sign of Q
^{63}Cu	29	$\frac{3}{2}$	$-ve$
^{65}Cu	29	$\frac{3}{2}$	$-ve$
^{69}Ga	31	$\frac{3}{2}$	$+ve$
^{71}Ga	31	$\frac{3}{2}$	$+ve$
^{75}As	33	$\frac{3}{2}$	$+ve$
^{79}Br	35	$\frac{3}{2}$	$+ve$
^{81}Br	35	$\frac{3}{2}$	$+ve$
^{85}Rb	37	$\frac{3}{2}$	unknown
^{87}Rb	37	$\frac{3}{2}$	unknown

A similar position arises in the region $8 < (N \text{ or } Z) < 16$, where the $1d_{5/2}$ and $2s_{1/2}$ levels compete. Here the shell of larger j , and predicted larger pairing energy, lies lowest, so that we do not expect anomalies. However, calculations not reproduced here indicate that, for forces of extended range, the interaction between $2s_{1/2}$ nucleons is definitely larger than that between $1d_{5/2}$ nucleons. This accounts for the unexpected spin of ^{19}F , the ground state configuration being mainly $(2s_{1/2})^3$. In ^{18}F , similarly, the calculations indicate that the ground and first excited (1 mev) states are $(s_{1/2})^2$ configurations, whereas the second excited (5.9 mev) state is the lowest of the $(d_{5/2})^2$ configurations. The remarkable lack of low excited states in this odd-odd nucleus, and the similar paucity in ^{18}O , are thus accounted for by the comparatively large interactions between $2s_{1/2}$ nucleons; that is, by the large pairing energy.

§ 3. THE DEPENDENCE ON A

We now investigate the dependence on A . Bohr and Wheeler (1939) find empirically that, averaging out over shells,

$$P = kA^{-3/4} \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad (5)$$

We take a suitably weighted average over different parts of the periodic table and suppose that the average pairing energy so obtained corresponds to the stable nucleus containing the average number of protons or neutrons. The results are indicated in the figure. The equations of the lines of best fit are :

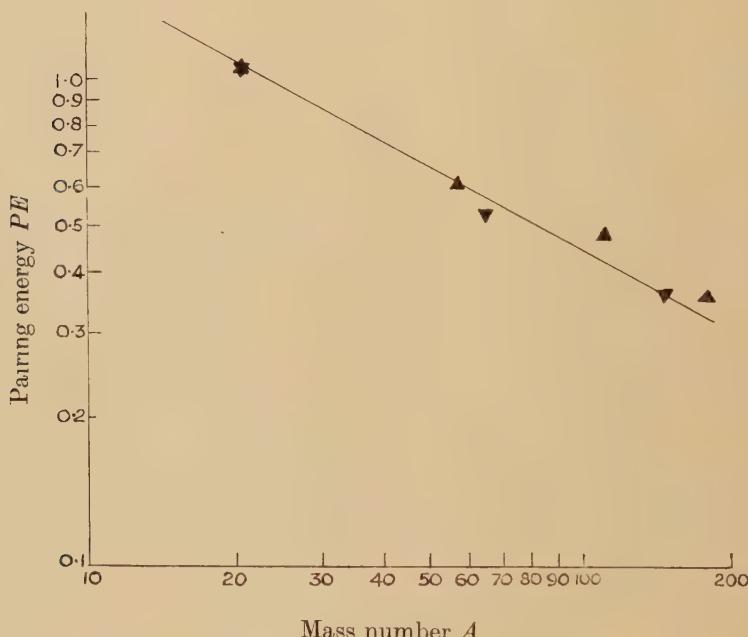
$$P_{\text{prot}} = kA^{-(0.55 \pm 0.05)}$$

$$P_{\text{neut}} = kA^{-(0.44 \pm 0.05)}$$

$$P_{\text{both}} = kA^{-(0.50 \pm 0.05)}.$$

The agreement with Bohr and Wheeler's semi-empirical formula is not good. These authors, however, do not explain how they arrive at their results. To be comparable with this, a semi-empirical formula must consider the complete periodic table, and not merely the fission product range with which Bohr and Wheeler are concerned.

The formula $P=c(2j+1)/A$ gives $P_{\text{both}}=kA^{-(0.60 \pm 1.0)}$. The difference between P_{prot} and P_{neut} was explicitly considered in Seaborg's (1952) semi-empirical mass formula for nuclei with $A > 210$, but this result cannot be used here since it is not extended over even one complete interval between two magic numbers. (He finds $P_{\text{prot}} > P_{\text{neut}}$, contradicting the theoretical result above.)



Averaged pairing energy as a function of A . ▲ Neutrons. ▼ Protons.

Since $P_{\text{neut}} < P_{\text{prot}}$, closed neutron shells should be more highly preferred than closed proton shells in spite of the Coulomb force, which tends to have the opposite effect. Table 4 shows that this is the case. In this table, the columns $(N)_M$, $(Z)_M$ denote the number of stable isotopes with N or Z equal to the closed shell number M . AvN , AvZ denote the average number of stable isotopes of constant N or Z in the region considered. $x = [(Z)_M/AvZ]/[(N)_M/AvN]$ is tabulated in the last column and is seen to be ≤ 1 in four cases out of five. Thus $P_{\text{neut}} > P_{\text{prot}}$ in general, as the theory requires.

Table 4. Number of Stable Isotopes with N or Z =a closed shell number M
 Notation explained in text. * α -emitters called stable.

M	Closed shell	$(N)_M$	$(Z)_M$	AvN	AvZ	x
38	$1f_{5/2}$	3	4	$3\frac{1}{4}$	6	0.7
50	$1g_{9/2}$	5	9	$3\frac{3}{4}$	$6\frac{3}{4}$	1.0
64	$2d_{5/2}$	4	7	$3\frac{1}{2}$	$6\frac{3}{4}$	0.9
76	$1h_{11/2}$	3	7	4	6	1.5
82*	$3s_{1/2}$	7	4	3	$6\frac{3}{4}$	0.3

§ 4. CONCLUSIONS

This detailed study of the values of the pairing energy integrals has enabled us to account for the following anomalies in the $j-j$ coupling model.

(i) The model now correctly allows for the possibility of the spin $\frac{3}{2}$ occurring for 77, 79 and 81 neutrons by predicting a large pairing energy in the $3s_{1/2}$ level. The spins $\frac{1}{2}$ occurring for 63, 65, 67 and 75 protons remain unexplained.

(ii) The signs of the quadrupole moments of odd proton nuclei in the $2p_{3/2}$ and $1f_{5/2}$ shells, which indicate a small pairing energy difference between these shells, is now explained.

(iii) The lack of low excited states in ^{18}F and ^{18}O , and the spin $\frac{1}{2}$ of ^{19}F are accounted for by the large pairing energy of the $2s_{1/2}$ nucleons.

(iv) General agreement with the observed dependence of pairing energy on A is not found, but it is correctly predicted that the pairing energy of protons is less than that of neutrons, in large nuclei of approximately the same mass.

This work has been assisted by the University of Chicago and by Trinity College, Cambridge. The author wishes also to place on record his gratitude to Professor M. G. Mayer who kindly checked the calculations and has been a constant source of encouragement and inspiration.

REFERENCES

BOHR, N., and WHEELER, J. A., 1939, *Phys. Rev.*, **56**, 426.
 MACK, J. E., 1950, *Rev. Mod. Phys.*, **22**, 64.
 MAYER, M. G., 1950, *Phys. Rev.*, **78**, 22.
 SEABORG, G. T., 1952, Reported at Pittsburgh Conference, 1952.

LXXX. *The Electrical Resistivity of Liquid Iron*

By R. W. POWELL

Physics Department, National Physical Laboratory, Teddington*

[Received March 13, 1953]

ABSTRACT

As a result of some preliminary experiments, which are described, the electrical resistivity of iron just above its melting point has been found to be 139 microhm cm²/cm. The resistivity of solid iron has also been measured and a value of 127.5 microhm cm²/cm is indicated at the melting point. Thus the increase in resistivity on fusion is only 9%, which is much less than would be expected on theoretical grounds and has been obtained for most other metals.

IN connection with theories relating to the Earth's interior, interest has been aroused (Elsasser 1946, 1950, Bullard 1948, 1949, 1950, Runcorn 1950) in the thermal and electrical conductivity of liquid iron. About forty years ago Bornemann and Wagenmann (1914) published values for the electrical resistivity of liquid iron over the range 1550° to 1650°C, and this appears to be the only experimental data available. Their measurements were actually made on three iron-carbon alloys, from which the values for pure iron were deduced. At the melting point, the value so derived for the resistivity of liquid iron agrees almost exactly with the value obtained for solid iron by linear extrapolation, from 1430°C of the writer's results for Armco iron (Powell 1939).

For most metals, however, the resistivity undergoes about a two-fold increase on fusion, and allowances for a change of this order had been made in connection with the above-mentioned geophysical theories. Indeed, if Mott's expression (1934)

$$\rho_l/\rho_s = (\nu_s/\nu_l)^2 = \exp (80L/T_M),$$

where ρ_l and ρ_s are the resistivities of the liquid and solid phases at the melting point, T_M °K, ν_l and ν_s the corresponding atomic frequencies and L the latent heat of fusion in kilo-joules per gram atom, is evaluated for iron, a ratio of 1.96 results.

In view both of this apparent disagreement with theory and of the geophysical requirements, further experimental work on the electrical resistivity of iron immediately above and below its melting point is being undertaken. The present note deals with some preliminary experiments which were made whilst 25 lb ingots of iron were being

* Communication from the National Physical Laboratory.

melted in the Metallurgy Division of the Laboratory. For the purpose of these measurements the ingot due to be melted was drilled with an axial hole about 1.5 in. in diameter. In this hole was supported a specially designed tube of alumina, about 15 in. long, 0.37 in. internal diameter and 1.45 in. external diameter. The dimensions of the retaining pot and ingot were such that about 5 in. of the alumina tube would ultimately become immersed in molten iron. Uniformly spaced in the centre of the thick walls of the tube were six 0.1 in. diameter holes. All were open at the top and were parallel to the axis. Two were close-ended and extended to depths of about 12 and 14 in. These were fitted with thermocouples. The lower ends of the other holes were open to the axial hole through small up-turned channels, the openings being at distances of approximately 0.5, 1.0, 4 and 4.5 in. from the bottom of the tube. These four holes were fitted with tungsten rods so that the extreme pair could serve as current leads and the inner pair as potential leads, when the circuit became completed by the ingot melting and the liquid iron filling the axial hole. The holes were up-turned so as to tend to trap any iron-tungsten alloy formed by the solution of the leads in the iron, thereby both reducing the rate of solution of the tungsten and the amount of contamination of the iron in the working section.

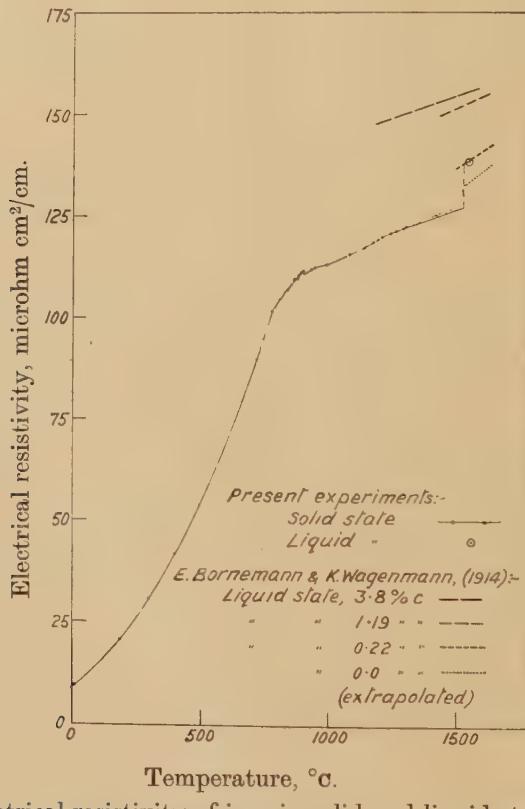
A preliminary calibration of the alumina tube had been carried out at normal temperatures using mercury. This enabled the effective value of the ratio 'area/length' to be determined and the value so obtained was increased by 1.5% to allow for the dimensional changes likely to occur between room temperature and 1550°C.

Two experiments were carried out in which the ingot was slowly heated in the normal manner in the Metallurgy Division's high frequency vacuum furnace. In the first the thermocouples used were of platinum and 13% rhodium platinum wires, and in the second assembly these were replaced by 5% rhodium platinum and 20% rhodium platinum wires. In both instances thermocouple troubles developed above 1500°C and these made it impossible to measure any temperatures when the iron had become molten and the resistivity observations could be commenced. From visual observation, however, the ingot could be seen to be molten and whilst the resistivity observations were being made, the temperature was estimated as within the range 1550°C to 1600°C. In the first experiment the resistivity circuit was maintained for about 50 minutes and the resistivity values obtained were 138.5, 140.0, 139.4, 137.3, 140.4 and 137.2 microhm cm²/cm. The power supply to the furnace had been switched off when the third value was obtained. This caused no marked change, nor is any definite variation noticeable in the course of the experiment. The mean of the six observations gives a value of 138.6 microhm cm²/cm.

The second experiment was carried out in much the same way, but the rate of heating was rather slower. This time the resistance circuit was retained for only about 25 minutes, but 18 sets of observations

were taken on the column of molten iron. The values ranged from 138.6 to 140.0 and the mean value was 139.1 microhm cm^2/cm . Thus the two experiments were in fair agreement, and indicate that the electrical resistivity of liquid iron close to its melting point is 139 microhm cm^2/cm .

After these experiments it was found that repeat measurements of the area/length ratios for the tubes agreed to within 0.5%, and that no cracks were apparent over the working section of the alumina tubes. In the case of the second tube this was also checked by means of a radiograph.



Electrical resistivity of iron in solid and liquid states.

In the accompanying figure is plotted all the information available at present for the electrical resistivity of liquid iron. A curve for the solid phase is also included. This has recently been obtained for a rod of high purity iron when heated under vacuum conditions in a platinum wound tubular furnace. The temperature measurements were obtained both from alumina sheathed platinum-platinum rhodium thermocouples and from similar thermocouples welded to the iron rod : these also served as potential leads. Up to about 1320°C there was agreement between

the thermocouples, but at higher temperatures those attached to the rod gave lower readings than the sheathed thermocouple. The readings of the latter were then used when plotting the results. At still higher temperatures there were indications that the sheathed thermocouple was also tending to give low readings so some allowance has been made for this by drawing the curve below the points in this region. No marked change in resistivity was apparent at the A4 point and the very slight amount of extrapolation necessary in this instance indicates that the resistivity of solid iron at the melting point is 127.5 microhm cm^2/cm . This leads to a value of only 1.09 for the ratio ρ_l/ρ_s .

In due course it is hoped to be able to continue the work using apparatus to be installed in the Physics Division, and if the difficulties of temperature measurement can be overcome, to extend the observations sufficiently far into the liquid phase to enable the temperature coefficient of the resistance of liquid iron to be determined.

Whilst the results so far obtained are regarded as of a preliminary nature, it is believed that they will be confirmed by this further work. In view of the interest which has been shown in the conductivity of liquid iron, publication at this early stage was decided upon.

ACKNOWLEDGMENTS

The author is indebted to Dr. E. C. Bullard, F.R.S. for suggesting the work on the resistivity of liquid iron and for his interest in its progress. He is also indebted to Dr. V. H. Stott, who was responsible for providing the specially constructed alumina tubes on which the success of the measurement depended, to Dr. N. P. Allen for permission to carry out the measurements on the liquid phase in the Metallurgy Division of the Laboratory and to Mr. G. C. H. Jenkins of that Division who operated the induction furnace and gave other assistance. Acknowledgment is also made to Mr. R. P. Tye and Mr. J. E. W. Jones of the Physics Division for their assistance in connection with the assemblies and observations on both the solid and liquid phases.

The work formed part of the general research programme of the National Physical Laboratory and is published with the approval of the Director.

REFERENCES

BORNEMANN, E., and WAGENMANN, K., 1914, *Ferrum*, **11**, 305.
 BULLARD, E. C., 1948, *Mon. Not. R. Ast. Soc., Geophys. Suppl.*, **5**, 248; 1949, *Proc. Roy. Soc. A*, **197**, 433; 1950, *Mon. Not. R. Ast. Soc., Geophys. Suppl.*, **6**, 36.
 ELSASSER, W. M., 1946 a, *Phys. Rev.*, **69**, 106; 1946 b, *Ibid.*, **70**, 212; 1950, *Rev. Mod. Phys.*, **22**, 1.
 MOTT, N. F., 1934, *Proc. Roy. Soc. A*, **146**, 465.
 POWELL, R. W., 1939, *Proc. Phys. Soc.*, **51**, 407.
 RUNCORN, S. K., 1950, *Nature, Lond.*, **166**, 974.

LXXXI. *Heat Conductivities of Superconductive Sn, In, Tl, Ta, Cb, and Al below 1°K*By K. MENDELSSOHN, F.R.S. and C. A. RENTON
Clarendon Laboratory, Oxford*

[Received April 13, 1953]

ABSTRACT

The heat conductivities of tin, indium, thallium, columbium, tantalum and aluminium below 1°K have been determined. At the lowest temperatures some of the specimens show a proportionality of the heat conductivity with T^3 and it has been suggested that this is heat transport by the crystal lattice only. The bearing of these results on the superconductive heat switch has been discussed.

§ 1. INTRODUCTION

In a previous communication (Olsen and Renton 1952) we have reported on a method of measuring heat conductivities below 1°K. In the arrangement used, the temperature gradient in the rod of the metal is measured at two intermediate points by carbon resistors. In this way the uncertainties inherent in experiments using magnetic thermometers at the ends of the specimen are avoided.

Measurements carried out with our method on a lead single crystal gave satisfactory and reproducible results and the work has now been extended to a number of other superconductors. While it is clear that the experiments carried out can only be considered as a preliminary survey of the field, a fairly clear pattern concerning the heat transport in superconductors at these very low temperatures emerges. A brief summary of the main results obtained to date is therefore given in this paper.

§ 2. THE SPECIMENS

The following metals were investigated :—

Tin, indium, thallium, tantalum, columbium and aluminium. Except for the polycrystalline tin rod Sn 2 and the aluminium sample Al 2, the specimens investigated were those measured at helium temperatures by Mendelssohn and Rosenberg (1952 a, b, 1953). A list of the samples used is given in the table.

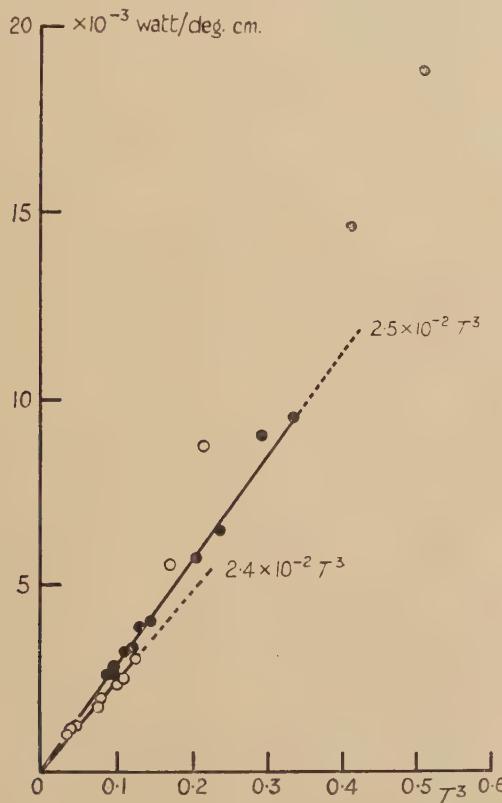
Specimen	State	Purity %
Sn 1	Single Crystal	99.997
Sn 2	Polycrystalline	99.997
In 2	Single Crystal	99.993
Tl 1	Polycrystalline	99.99
Al 2	Polycrystalline	unknown
Ta 1	Polycrystalline	99.98
Cb 1	Polycrystalline	99.99

* Communicated by the Authors.

§ 3. RESULTS

In fig. 1 the thermal conductivity K_s of the tin single crystal Sn 1 and the indium single crystal In 2 is plotted against the cube of the absolute temperature. At the lowest temperatures the heat conductivity in both cases is proportional to T^3 . At temperatures of $\sim 0.5^\circ\text{K}$ for Sn 1 and of $\sim 0.7^\circ\text{K}$ for In 2, K_s begins to rise more rapidly. This is exactly the same behaviour as was found in a lead single crystal (Olsen and Renton 1952) which showed departure from a T^3 function at $\sim 0.9^\circ\text{K}$.

Fig. 1

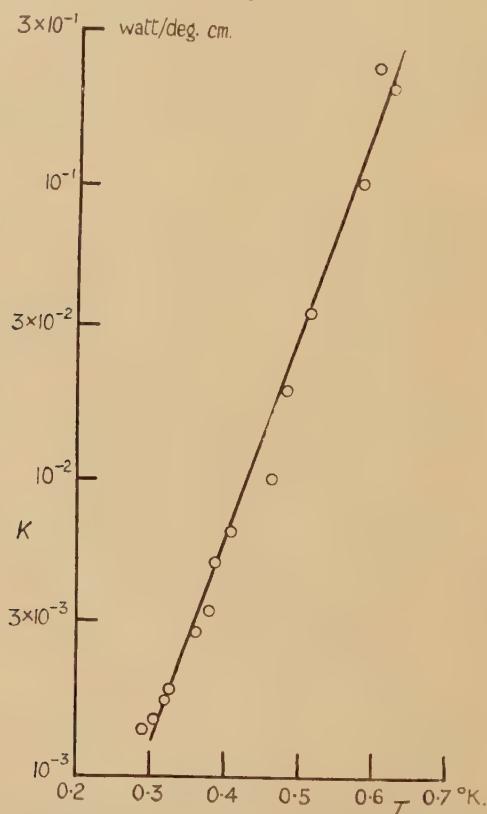


Heat conductivity of single crystals of tin \circ and indium \bullet below 1°K plotted against T^3 .

The heat conductivity of the polycrystalline thallium rod Tl 1 is plotted logarithmically against the absolute temperature in fig. 2. While a certain amount of scatter has to be admitted, it seems that the results over the whole temperature range can at a first approximation be represented by a straight line, indicating an exponential rise of K_s with temperature. There is possibly an indication of a slower rise at the low temperature end.

The results on Sn 2, Ta 1, Cb 1, and Al 2 are shown in a log-log plot in fig. 3. It is significant that two different experiments on Sn 2 yielded slightly different results. Moreover the power law at the low temperature end gives a rise of thermal conductivity with $\sim T^{1.3}$. These facts indicate that on demagnetization some transverse magnetic flux was 'frozen in' and that this amount was different in the two experiments. It is to be noted that no such effect occurred in the measurements on Sn 1, In 2 and Tl 1 which are all the results of more than one experiment.

Fig. 2

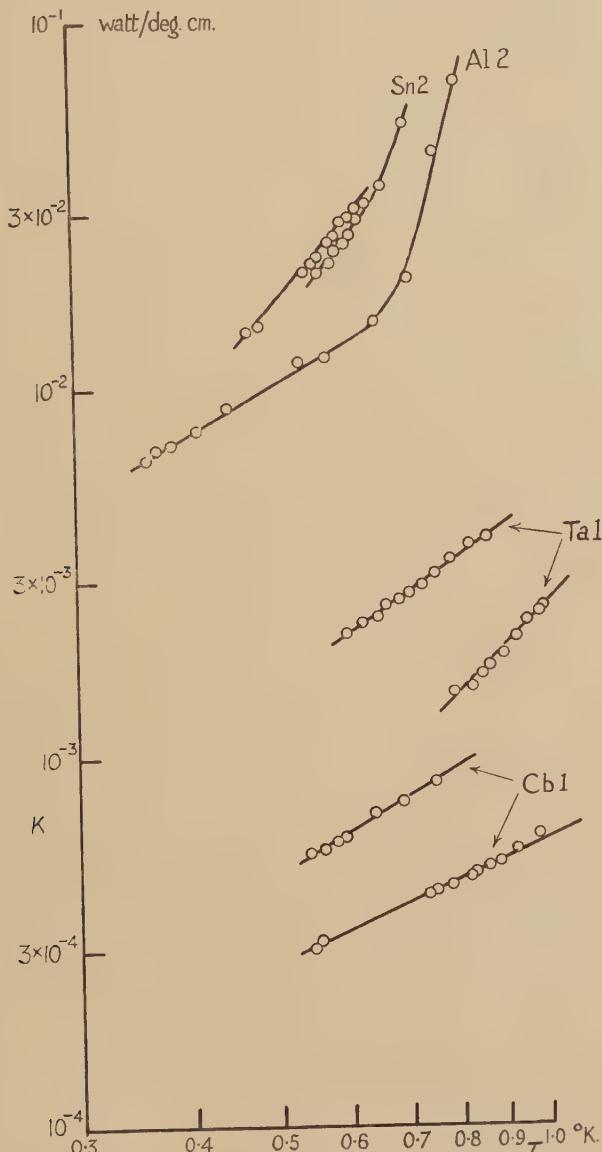


Heat conductivity of polycrystalline thallium.

The effect of 'frozen in' magnetic fields is enhanced in the case of Ta 1 and Cb 1. The tendency of 'hard' superconductors to form a magnetic 'sponge' (Mendelssohn 1935) similar to the behaviour of alloys is well known and the present results are therefore not surprising. It is interesting to note that in all four experiments the power law is smaller than 3 and larger than 1, indicating an intermediate stage between the T^3 law to be expected for a pure superconductor and direct proportionality to absolute temperature, characteristic of a normal metal.

Only one series of experiments was carried out on Al 2 because similar to Ta 1 and Cb 1 the physical purity of the sample was not very satisfactory. Here also the results indicate the presence of 'frozen in' flux which is probably responsible for the variation of the heat conductivity with

Fig. 3



Heat conductivity of polycrystalline tin, tantalum, columbium and aluminium.

$\sim T^{1.4}$ below $\sim 0.6^\circ\text{K}$. However, above this temperature the curve rises much more steeply, an effect which is akin to that observed in Sn 1, In 2 and the lead single crystal.

§ 4. DISCUSSION

Our results indicate that the cases of Sn 1, In 2 and the lead single crystal investigated earlier are those characteristic of a pure superconducting metal. In all three instances single crystals of very high purity were at our disposal. The relatively low melting points of these metals, moreover, make it likely that the crystals are relatively unstrained. It is therefore satisfying to note that the pattern of behaviour is the same in the three metals. There is a good adherence to a T^3 law for the thermal conductivity at low temperatures followed by a steeper, possibly exponential, rise. From earlier work it is known that as absolute zero is approached all the conduction electrons normally contributing to the entropy are passing into the superconductive state (Daunt, Horseman and Mendelssohn 1939) and that the entropy of the system of superconductive electrons itself, even at finite temperatures, is zero (Daunt and Mendelssohn 1946). It is therefore reasonable to assume that the heat conductivity at the lowest temperatures, represented by a T^3 function, is that of the crystal lattice alone. Comparing our results in this region with Casimir's (1938) calculation, the numerical values of the observed heat conductivities are between five and ten times smaller than those theoretically predicted for an ideal single crystal. This discrepancy might be accounted for by the existence of scattering centres in our specimens. It is, of course, possible that our values hide an electronic heat conductivity term also proportional to T^3 . However, considering that our results are already several times smaller than the ideal lattice conduction such an electronic term must be very small.

The exact function of the rise in K_s following the T^3 law cannot as yet be determined for these single crystals because values in the temperature range between 1° and 2°K , which requires a different technique, would be needed. However, an indication is provided by the exponential function observed in Tl 1. Such an exponential law has indeed been suggested by Koppe (1947) for the lowest region of the electronic heat conductivity. Assuming then that the T^3 region is due to lattice conductivity and that the electronic conductivity follows an exponential law, one might expect the change over from the T^3 function to the exponential to occur at a higher relative temperature for a more ideal specimen. Expressing the temperature at which this change over occurs as a fraction in a reduced scale, T/T_c , where T_c is the transition temperature, we obtain for Sn 1 : ~ 0.15 , for In 2 : ~ 0.2 and for the lead single crystal : ~ 0.15 . The lowest reduced temperature to which Tl 1 has been measured is 0.13 with as yet no clear change over to a T^3 law. This is in keeping with the fact that Tl 1 is polycrystalline.

As in the case of lead, magnetic cycles between the superconductive and normal state were carried out for Sn 2, In 2 and Tl 1. Mendelssohn and Olsen (1950 a, b), noted the peculiar effect of a maximum in the heat resistance in the intermediate state of superconductors which was also found with the lead single crystal. Detwiler and Fairbank (1952) confirmed this effect in

observations on tin and indium and suggested it might be a property of all pure superconductors. It is interesting in this respect that indeed such a maximum was found for Sn 2 at 0.43°K, but was absent in In 2 at 0.60°K and Tl 1 at 0.32°K.

§ 5. THE SUPERCONDUCTIVE HEAT SWITCH

The change in thermal conductivity between the superconductive and the normal state is important for the operation of a 'heat switch' in the magnetic temperature range. As was already pointed out in the case of our earlier measurements (Olsen and Mendelsohn 1950 a, Olsen and Renton 1952) and also in the work of Heer and Daunt (1949), the ratio of normal to superconductive heat conduction K_n/K_s becomes very high at these low temperatures. In fact, using an ordinary heat conductivity measuring arrangement it is impossible to determine with any degree of accuracy both K_n and K_s because of the big difference in absolute values. However, since for most of our specimens it has been found in the helium range that K_n varies proportionally to the absolute temperature, extrapolation can be made with a fair degree of confidence. The ratio K_n/K_s at the lowest temperatures can therefore be expressed as aT/bT^3 or αT^{-2} . Values of the coefficient α as estimated from our experimental results are as follows: for Sn 1 : 350, for In 2 : 115, and for the lead single crystal : 100. Care has to be taken against any appreciable 'freezing in' of normal material in the 'open' position of the switch which, as is shown by the examples of fig. 3, will increase K_s .

REFERENCES

CASIMIR, H. B. G., 1938, *Physica*, **5**, 595.
 DAUNT, J. G., HORSEMAN, A., and MENDELSSOHN, K., 1939, *Phil. Mag.*, **27**, 754.
 DAUNT, J. G., and MENDELSSOHN, K., 1946, *Proc. Roy. Soc. A*, **185**, 225.
 DETWILER, D. P., and FAIRBANK, H. A., 1952, *Phys. Rev.*, **86**, 574.
 HEER, C. V., and DAUNT, J. G., 1949, *Phys. Rev.*, **76**, 854.
 KOPPE, H., 1947, *Ann. Phys.*, **1**, 405.
 MENDELSSOHN, K., 1935, *Proc. Roy. Soc. A*, **152**, 34.
 MENDELSSOHN, K., and OLSEN, J. L., 1950 a, *Proc. Phys. Soc. A*, **63**, 2 ; 1950 b, *Phys. Rev.*, **80**, 859.
 MENDELSSOHN, K., and ROSENBERG, H. M., 1952 a, *Proc. Phys. Soc. A*, **65**, 385 ; 1952 b, *Ibid.*, **65**, 388 ; 1953, *Proc. Roy. Soc. A*, in the press.
 OLSEN, J. L., and RENTON, C. A., 1952, *Phil. Mag.*, **43**, 946.

LXXXII. CORRESPONDENCE

On the Mutual Transformation of Lattices

By B. A. BILBY

Royal Society Sorby Research Fellow, The University of Sheffield

[Received May 18, 1953]

§ 1. GENERATING NODES BETWEEN TWO LATTICES

In view of the accompanying note (Basinski and Christian 1953) a brief account is given below of some unpublished investigations previously reported only in outline (Bilby 1953); the results here given are in fact rather more general.

A 'pole mechanism' has been proposed (Cottrell and Bilby 1951) to explain the production of macroscopically homogeneous deformations by dislocation movements. In this three dislocation lines 1, 2, 3 with Burgers vectors \mathbf{b}_1 , \mathbf{b}_2 , \mathbf{b}_3 meet at a node, and one, say \mathbf{b}_3 , rotates about the node in a plane whose normal is \mathbf{k} . The lines 1 and 2 lie on opposite sides of the plane \mathbf{k} and form the 'pole'. If all the dislocation lines lie in one lattice and the line 3 is perfect, then we must have $(\mathbf{b}_3 \cdot \mathbf{k})=0$. The arrangement is thus a Frank-Read source if $(\mathbf{b}_2 \cdot \mathbf{k})=-(\mathbf{b}_1 \cdot \mathbf{k})=0$; otherwise successive planes $(\mathbf{b}_2 \cdot \mathbf{k})=-(\mathbf{b}_1 \cdot \mathbf{k})$ apart suffer a relative displacement \mathbf{b}_3 and a mean macroscopic shear of $|\mathbf{b}_3|(\mathbf{b}_2 \cdot \mathbf{k})$ is produced. This is 'homogeneous slip'. The lines 1 and 2 may however lie in different lattices, the plane \mathbf{k} forming the boundary between them. Then if the Burgers vectors and the lattices are suitably related, rotation of the line 3 generates one lattice from the other. We call such an arrangement of dislocation lines a *generating node between the two lattices*. The concept has been used in theories of the twinning of iron (Cottrell and Bilby 1951) and of cadmium (Millard and Thompson 1952).

Any homogeneous deformation carrying the vector \mathbf{r} to \mathbf{r}' ,

$$\mathbf{r} \rightarrow \mathbf{r}' = \mathbf{r} \cdot \Phi + \mathbf{t}$$

where Φ is a constant dyadic and \mathbf{t} a constant vector, converts a Bravais lattice with vectors $\mathbf{C}(m)=m^i \mathbf{c}_i$ to another lattice translated \mathbf{t} with respect to the first. For the vectors $\mathbf{C}'(m)-\mathbf{t}=m^i(\mathbf{c}_i \cdot \Phi)$ are those of a lattice $\mathbf{P}(m)=m^i \mathbf{p}_i$ with basis $\mathbf{p}_i=(\mathbf{c}_i \cdot \Phi)$. As a result of the deformation any vector $\mathbf{C}(m)$ of the C lattice becomes some vector $\mathbf{P}(n)$ of the P lattice: we say that $\mathbf{C}(m)$ generates $\mathbf{P}(n)$. When \mathbf{p}_i is the basis for the P lattice the generated vector $\mathbf{P}(n)$ has the same numerical components as the vector $\mathbf{C}(m)$ generating it, that is, $m^i=n^i$. On the other hand, given two Bravais lattices there exists an infinite set of deformations Φ carrying the one into the other. In the present discussion it is assumed that a definite Φ has been chosen using some further criterion, for example, that of least displacement of the lattice points.

Given the C lattice it is not possible to generate an arbitrary P lattice from it by a simple pole mechanism. The translation \mathbf{t} requires special discussion and we assume first that $\mathbf{t}=0$. Take the origin in the boundary and assume that the dislocation 3 climbs along dislocation 1 into the C lattice. Then the displacement of a general vector \mathbf{r} is

$$\mathbf{r}' - \mathbf{r} = \frac{(\mathbf{r} \cdot \mathbf{k})}{(\mathbf{b}_1 \cdot \mathbf{k})} \mathbf{b}_3$$

and if there is no vector $\mathbf{C}(n)$ such that $|(\mathbf{C}(n) \cdot \mathbf{k})| < |(\mathbf{b}_1 \cdot \mathbf{k})|$ the deformation is homogeneous, each lattice point moving in a direction \mathbf{b}_3 an amount proportional to its distance from the boundary plane. Although the dislocation 3 moves in the plane \mathbf{k} , its Burgers vector \mathbf{b}_3 may be arbitrary. This is because the usual kinematic restriction on the glide motion of a dislocation, namely that such motion can take place only in planes containing its Burgers vector and its line, is relaxed, since the dislocation here lies at the boundary between two lattices, and the conversion of one to the other may involve contraction or expansion normal to the boundary. Thus, if we take the unit vector \mathbf{i} perpendicular to \mathbf{k} and write $\mathbf{b}_3/(\mathbf{b}_1 \cdot \mathbf{k}) = \lambda \mathbf{i} + \mu \mathbf{k}$, the most general homogeneous deformation Φ which can be produced by the pole mechanism is

$$\Phi = \mathbf{I} + \mathbf{k}(\lambda \mathbf{i} + \mu \mathbf{k}) \quad \dots \quad (1)$$

This, of course, is the most general homogeneous distortion leaving one plane (normal to \mathbf{k}) undistorted and unrotated. When $\mu=0$, Φ represents simple shear, and when $\lambda=0$, uniaxial strain along the \mathbf{k} direction.

Given two lattices related by such a deformation we may always construct a generating node between them in the following way. The difference between the vectors $\mathbf{P}(n)$ and $\mathbf{C}(m)$ is

$$\mathbf{C}(n) + (\mathbf{C}(n) \cdot \mathbf{k})(\lambda \mathbf{i} + \mu \mathbf{k}) - \mathbf{C}(m)$$

so that dislocation lines with Burgers vectors $-\mathbf{P}(n) - \mathbf{C}(m)$, $\mathbf{C}(n) - \mathbf{C}(m)$ and $(\mathbf{C}(n) \cdot \mathbf{k})(\lambda \mathbf{i} + \mu \mathbf{k})$ can form a node. Clearly the line of Burgers vector $\mathbf{C}(m)$ is not essential and the node becomes :

$$-\mathbf{P}(n) + \mathbf{C}(n) + (\mathbf{C}(n) \cdot \mathbf{k})(\lambda \mathbf{i} + \mu \mathbf{k}) = 0 \quad \dots \quad (2)$$

It is readily verified that if $-\mathbf{P}(n)$ is the pole in P , $\mathbf{C}(n)$ the pole in C , and $(\mathbf{C}(n) \cdot \mathbf{k})(\lambda \mathbf{i} + \mu \mathbf{k})$ the sweeping dislocation, then this is a generating node between the two lattices. The vector $\mathbf{C}(n)$ generates $\mathbf{P}(n)$ and $(\mathbf{C}(n) \cdot \mathbf{k})(\lambda \mathbf{i} + \mu \mathbf{k})$ is the difference between them. By analogy with the twinning dislocation (Frank 1951) we call the last dislocation a *transformation* dislocation.

It is not in fact necessary that the vectors $\mathbf{P}(n)$ and $\mathbf{C}(n)$ be perfect lattice vectors provided that $\mathbf{C}(n)$ generates $\mathbf{P}(n)$ and that the vectors are suitable fault vectors of the C and P lattices respectively. It is also not necessary that the plane \mathbf{k} be rational in either lattice: it is defined by being the undistorted unrotated plane of the deformation. It is, however, important that the $-\mathbf{P}(n)$ dislocation lies in P and the $\mathbf{C}(n)$ dislocation in C . For otherwise the transformation dislocation tends to climb back into

the P lattice as it moves to produce the displacements generating that lattice. As these will not generally be permitted in the P lattice, the node is sessile and may be called *anti-generating*. It is easy to see that any generating node between the P and C lattices is essentially of type (2). For dislocation lines $\mathbf{P}(m)$ in the P lattice, $\mathbf{C}(n)$ in the C lattice and $(\mathbf{C}(n) \cdot \mathbf{k})(\lambda\mathbf{i} + \mu\mathbf{k})$ in the boundary must form a node: thus

$$[\mathbf{C}(n) + \mathbf{C}(m)] \cdot [\mathbf{I} + \mathbf{k}(\lambda\mathbf{i} + \mu\mathbf{k})] = 0.$$

Now $\det[\mathbf{I} + \mathbf{k}(\lambda\mathbf{i} + \mu\mathbf{k})] = (1 + \mu)$, and since $\mu \neq -1$ ($\mu = -1$ corresponds to 100% compression along \mathbf{k}), the determinant does not vanish and so $n^i = -m^i$, and the $\mathbf{C}(n)$ vector must generate the vector $-\mathbf{P}(m)$.

We have now to show how the translation \mathbf{t} may be produced. This is achieved if two parallel dislocation lines of equal and opposite Burgers vectors $\pm \mathbf{t}$ are associated with the transformation dislocation and move with it. These lines lie a distance $(\mathbf{C}(n) \cdot \mathbf{k})$ apart in the \mathbf{k} direction, which is the height of the step in the boundary corresponding to the transformation dislocation. Taken together these additional dislocation lines have zero Burgers vector and their stress field is that of a dislocation doublet. They may, for example, represent a moving line of dilatation or compression.

§ 2. THE TRANSFORMATION OF COBALT

The preceding discussion has dealt with the mutual transformation of lattices, by homogeneous deformation. In most examples, however, the crystals transforming have structures with a basis, or even if this is not so, the deformation is not a single homogenous one. We must therefore envisage inhomogeneous atom movements. When these involve the mutual translation of atoms on Bravais lattices undergoing the same homogeneous deformation, a formal description can be given in the way that the translation \mathbf{t} was described above. The inhomogeneous movements in the cobalt transformation are, however, of a very simple kind. If we refer the face-centred cubic lattice to a cell \mathbf{c}_i (with reciprocal cell \mathbf{d}^i) related to the usual cubic axes \mathbf{a}_i by

$$\mathbf{c}_i = \alpha_i \mathbf{a}_k \quad \alpha_i = \frac{1}{2} \begin{bmatrix} 1 & 0 & \bar{1} \\ 1 & 1 & \bar{2} \\ 1 & 1 & 2 \end{bmatrix}$$

the structure basis is $[000]$ and $[0 \frac{1}{2} \frac{1}{2}]$. A hexagonal structure of axial ratio s can now be generated by a homogeneous deformation

$$\Phi = \mathbf{I} + (1/2\sqrt{6})[\mathbf{d}^3][s\mathbf{c}_2 + (3s - 2\sqrt{6})\mathbf{c}_3]$$

of the $[000]$ lattice, while the $[0 \frac{1}{2} \frac{1}{2}]$ lattice suffers a similar deformation, together with a translation $-(\frac{1}{6})\mathbf{c}_2$. Vectors of the type

$$\mathbf{v} = [n_1, n_2 + \frac{1}{2}, n_3 + \frac{1}{2}] - [n_1, n_2, n_3],$$

change by $t((\frac{1}{3})\mathbf{c}_2 + \mathbf{c}_3)$ where $t = (3s - 2\sqrt{6})/4\sqrt{6}$. Thus a generating node between the [000] lattices based on Φ , in which the pole dislocation $\mathbf{C}(m)$ has $(\mathbf{C}(m) \cdot \mathbf{d}^3) = 1$, which produces no change in \mathbf{v} , and a pair of dislocations with Burgers vectors $\pm t((\frac{1}{3})\mathbf{c}_2 + \mathbf{c}_3)$ associated with the sweeping dislocation, will produce the required transformation.

The simplest node is

$$-\mathbf{P}[001] + \mathbf{C}[001] + (1/2\sqrt{6})[0, s, 3s - 2\sqrt{6}]_c = 0,$$

or in the conventional axes of the hexagonal and cubic phases :

$$-[00\cdot 1]_H + \frac{1}{2}[112]_a + (1/2\sqrt{6})[2s - \sqrt{6}, 2s - \sqrt{6}, 2(s - \sqrt{6})]_a = 0.$$

With the $\frac{1}{2}[112]_a$ dislocation dissociated into the dislocations

$$\frac{1}{2}[011]_a + \frac{1}{2}[101]_a,$$

this is a node of the kind considered in the accompanying note (Basinski and Christian 1953), except that we allow for the slight contraction perpendicular to $\{111\}_a$ which occurs during the transformation.

REFERENCES

BASINSKI, Z. S., and CHRISTIAN, J. W., 1953, *Phil. Mag.*, **44**, 791.
 BILBY, B. A., 1953, *Year Book of the Royal Society*, p. 217.
 COTTRELL, A. H., and BILBY, B. A., 1951, *Phil. Mag.*, **42**, 573.
 FRANK, F. C., 1951, *Phil. Mag.*, **42**, 809.
 MILLARD, D. J., and THOMPSON, N. F., 1952, *Phil. Mag.*, **43**, 422.

On Inverse Perturbation

By A. CUNLIFFE and R. N. GOULD
 University College of Hull

[Received May 18, 1953]

PERTURBATION theory as normally used gives the perturbed eigenvalues and eigenfunctions when the perturbed operator is known. Experiment, however, usually gives the changes in the eigenvalues brought about by the perturbation. It therefore seems desirable to carry out an inverse process where the perturbing operator and perturbed eigenfunctions are expressed in terms of one or more of the perturbed eigenvalues. The present note sets out conditions for which this inverse process is practicable.

Let W_n, ψ_n be the n th eigenvalues and eigenfunctions of the unperturbed equation

$$H\psi = W\psi, \quad \dots \dots \dots \dots \dots \quad (1)$$

where H is the unperturbed operator. Suppose that W_n , ψ_n become W_{np} , ψ_{np} when a perturbation U is added to H so that the perturbed equation is

$$(H+U)\psi = W\psi. \quad \dots \dots \dots \quad (2)$$

It will be supposed that (1) and (2) hold throughout a region τ which is the same for both the unperturbed and the perturbed systems, and that the unperturbed eigenfunctions form a complete set. Now choose $\lambda = (W_{np} - W_n)/W_n$ as a parameter in the Rayleigh-Schrödinger method (Mott and Sneddon 1948). Then if U can be expressed in the form

$$U = a_1\theta_1 + a_2\theta_2 + \dots + a_k\theta_k, \quad \dots \dots \dots \quad (3)$$

where the θ 's are known operators but the a 's are unknown constants independent of position, it is possible to calculate approximate values for the a 's if k different λ 's (corresponding to different states n) are known.

Thus if the unperturbed system is *non-degenerate*, and

$$\lambda_1 = \frac{W_{1p} - W_1}{W_1}, \quad \lambda_2 = \frac{W_{2p} - W_2}{W_2}, \quad \dots \dots \quad \lambda_k = \frac{W_{kp} - W_k}{W_k},$$

$$a_s = \left| \begin{array}{c} T_{11}T_{21} \dots T_{(s-1)1}W_1\lambda_1T_{(s+1)1} \dots T_{k1} \\ T_{12}T_{22} \dots T_{(s-1)2}W_2\lambda_2T_{(s+1)2} \dots T_{k2} \\ \dots \dots \dots \\ T_{1k}T_{2k} \dots T_{(s-1)k}W_k\lambda_kT_{(s+1)k} \dots T_{kk} \end{array} \right| \left/ \left| \begin{array}{c} T_{11}T_{21} \dots T_{k1} \\ T_{12}T_{22} \dots T_{k2} \\ \dots \dots \dots \\ T_{1k}T_{2k} \dots T_{kk} \end{array} \right| \right. \quad \dots \dots \dots \quad (4)$$

where

$$T_{ij} = \int \psi_j^* \theta_i \psi_j d\tau.$$

Equation (4) is accurate to first order in the λ 's. To this same order of accuracy,

$$\psi_{np} = \psi_n + \sum_j \frac{\psi_j^* U \psi_n d\tau}{W_n - W_j} \psi_j, \quad \dots \dots \dots \quad (5)$$

where U in (5) is obtained from (3) by substituting for the a 's from (4) and the summation \sum_j excludes $j=n$. In theory it is possible to extend the process so as to obtain equations corresponding to (4) and (5) which are of second or higher order accuracy in the λ 's.

On the other hand, λ 's arising from unperturbed *degenerate* states can be used in the calculation of the a 's, but this will give rise to ambiguity in some cases.

As an example of the use of inverse perturbation, consider an electromagnetic cavity resonator having a vacuum dielectric and perfectly conducting completely enclosing walls. Let the cavity be of such a shape that the eigenvalues γ_n and eigenfunctions E_n satisfying the unperturbed equation

$$\nabla^2 E = \gamma E \quad \dots \dots \dots \quad (6)$$

within the cavity, and

$$E \times dS = 0$$

at the cavity wall S , can be calculated. If the cavity is perturbed by introducing dielectric material of dielectric constant ϵ inside it, then (6) must be replaced by the perturbed equation

$$\nabla^2 E = \epsilon \gamma E. \quad \dots \dots \dots \dots \dots \quad (7)$$

Suppose that ϵ can be written in the form

$$(\epsilon - 1) = b_1 \theta_1 + b_2 \theta_2 + \dots \dots \dots \dots \quad (8)$$

where the θ 's are known functions of position but the b 's are unknown constants. The inverse perturbation method can then be used to find the b 's. If ϵ is a completely unknown function of position, the expansion (8) would involve an infinite number of the b 's where the θ 's form a complete set of scalar functions. To find the b 's would then be impracticable. An approximate problem where there are only a finite number of the b 's must therefore be solved. In practice, ϵ will be frequency dependent. Consequently, unless this frequency dependence is known, it will be necessary to take perturbed eigenvalues which lie within a reasonably narrow waveband. In the particular case where $(\epsilon - 1) = b_1 \theta_1$, application of the foregoing theory gives for non-degenerate modes,

$$(\epsilon - 1) = \left[-\frac{1}{I_{nn}} \lambda + \left(\frac{1}{I_{nn}} + \frac{\gamma_n}{I_{nn}^3} \sum_j \frac{I_{nj}^2}{\gamma_n - \gamma_j} \right) \lambda^2 + \dots \right] \theta_1 \quad \dots \quad (9)$$

where $\lambda = (\gamma_{np} - \gamma_n)/\gamma_n$, $I_{nn} = \int \theta_1 E_n^2 d\tau$, $I_{nj} = \int \theta_1 E_n \cdot E_j d\tau$, the volume integrals being throughout the whole volume τ of the cavity. If the perturbation is brought about by the introduction of a piece of uniform material of dielectric constant ϵ , by putting $\theta_1 = 1$ in the region of the material and $\theta_1 = 0$ elsewhere, (9) gives the dielectric constant of the inserted material

$$\epsilon = 1 - \frac{\lambda}{\int_D E_n^2 d\tau} + \left[\frac{1}{\int_D E_n^2 d\tau} + \frac{\gamma_n}{\left(\int_D E_n^2 d\tau \right)^3} \sum_j \frac{\left(\int_D E_n \cdot E_j d\tau \right)^2}{\gamma_n - \gamma_j} \right] \lambda^2 \dots \quad \dots \quad (10)$$

where the volume integrals $\int_D d\tau$ are throughout the dielectric region only. If (10) is written to first order in λ , the equation reduces to that given by Bethe and Schwinger (1943) for the particular case where the dielectric has its surface parallel to the electric field lines.

The authors wish to thank Professor L. S. Palmer for his encouragement and Mr. J. M. Hough for his interest and for very valuable discussions.

REFERENCES

BETHE and SCHWINGER, 1943, *Perturbation Theory for Cavities*, M.I.T. Report D1-11-7, 4/3/1943.
 MOTT and SNEDDON, 1948, *Wave Mechanics and its Applications* (Oxford), p. 73.

A Note on the Adsorption of Helium on Glass

By EARL LONG and LOTHAR MEYER

Institute for the Study of Metals, The University of Chicago, U.S.A.

[Received May 27, 1953]

BREWER AND MENDELSSOHN (1953) have recently published data on the adsorption of helium on glass at a number of temperatures below the boiling point, and at saturations ranging from 30% to 95%. The data are presented by plotting the amount adsorbed at constant saturation P/P_0 as a function of temperature, for a series of saturations. The authors note that the quantity adsorbed *increases* with temperature in the He-II region until, at the λ -point, a sharp break occurs and the quantity adsorbed then decreases with increasing temperature in the He-I region, the results being in disagreement with the earlier work of Frederikse and Gorter (1950) on steel and on iron oxide, Long and Meyer (1949) on iron oxide, Strauss (1952) on iron oxide, and the later data of R. Bowers (1953) on aluminium.

In terms of a family of the usual adsorption isotherms, these results mean that below T_λ , and at constant amount adsorbed, the saturation P/P_0 *decreases* with increasing temperature. This does not seem to fit into the thermodynamic treatment of adsorption systems, since we have

$$RT^2 \frac{d}{dT} \log P/P_0 = \overline{\Delta H_f} - \Delta H_0 \dots \dots \quad (1)$$

where $\overline{\Delta H_f}$ is the differential molal heat of adsorption and ΔH_0 is the heat of vaporization of bulk liquid. A negative sign for the term involving P/P_0 thus requires that the heat of adsorption be less than ΔH_0 for the bulk liquid. This furthermore leads to impossibly high values for the differential entropy of the adsorbed film (cf. Long and Meyer 1953, eqn. 7).

The break at the λ -point is equally difficult to reconcile with the requirements of thermodynamics. Again, such a break means that at constant amount adsorbed the equilibrium pressure P must change rapidly within an exceedingly small temperature range. The Clausius-Clapeyron equation then requires an extremely high heat of adsorption in this narrow temperature range. However, the thermodynamic limitation is rather rigorous, since

$$d \frac{\Delta H}{dT} = \Delta \frac{dH}{dT} = \Delta C \dots \dots \dots \quad (2)$$

Near the λ -point, the heat capacity of the vapour still exceeds that of the adsorbed phase; consequently, the maximum possible value for the right side of eqn. (2) is the heat capacity of the gas, 5 cal/mole/deg. From fig. 1 of Brewer and Mendelssohn it appears that the break occurs over a

temperature range of the order of magnitude of 0.01° . Therefore, a maximum change in the heat of adsorption is ~ 0.05 cal/mole, whereas the heat itself cannot be less than 22 to 25 cal/mole. Such a small change would produce a practically imperceptible change in the slope of P vs T , a result which is indeed confirmed by the experiments of Long, Meyer, and Strauss (cf. Long and Meyer 1953, p. 25).

REFERENCES

BOWERS, R., 1953, *Phil. Mag.*, **44**, 467, 485.
 BREWER, D. F., and MENDELSSOHN, K., 1953, *Phil. Mag.*, **44**, 340, 559.
 FREDERIKSE, H. P. R., and GORTER, C. J., 1950, *Physica*, **16**, 403.
 LONG, E., and MEYER, L., 1949, *Phys. Rev.*, **76**, 440; 1953, *Advances in Physics*, **2**, 1.
 STRAUSS, A. J., 1952, *Thesis*, Chicago. (See Long and Meyer 1953.)

Inconsistencies in Adsorption Experiments of Helium II

By D. F. BREWER and K. MENDELSSOHN, F.R.S.
 The Clarendon Laboratory, Oxford

[Received May 27, 1953]

THE interesting comments by Long and Meyer (1953 b) serve to emphasize the point which we tried to make in our original paper (Brewer and Mendelssohn 1953 a), namely that our results are incompatible with the simple concept of an adsorbed film. Unless our observations are in themselves erroneous, for which we cannot at present see any reason, the only alternative is to assume the existence in adsorption measurements below the lambda point of a factor of which no account had been taken so far. It was for this reason that we suggested the presence of bulk liquid below the lambda point at pressures much less than the saturation pressure. Following up this idea, we postulated an anomalous surface tension effect, the existence of which we have since demonstrated (Brewer and Mendelssohn 1953 b). Further evidence for the existence of this new effect is provided by dielectric measurements recently carried out at this laboratory by Hatton, Rollin and Seymour (to be published shortly).

The disagreement between our results and other adsorption measurements had already been stressed in our first paper (1953 a) but here the position is even worse than stated by Long and Meyer. In 1949, Long and Meyer reported adsorption measurements on Fe_2O_3 in which, for constant amount adsorbed, the saturation p/p_0 was *constant* within 2% between 1.53° and 2.11°K . In 1950, the same authors, using a closed system, say that "the saturation stayed substantially constant up to the

lambda point, then changed practically discontinuously when passing the lambda point". On the other hand, the adsorption measurements of Strauss (Strauss 1952, Long and Meyer 1953 a), also on Fe_2O_3 , show an increase of saturation at constant amount adsorbed below the lambda point, and it is stated (Long and Meyer 1953 a), that in closed vessel experiments, "the only unusual effects observed were slight discontinuities in the plots at the bulk liquid lambda-point, due to small errors in the hydrostatic corrections for the bath temperature above 2.186°K ". However, the $\log p - 1/T$ plot made in this paper again differs somewhat from one given earlier (Strauss, Meyer and Long 1951). Finally the adsorption measurements of Frederikse and Gorter (1950) also on Fe_2O_3 , disagree with those of Long and Meyer (1949) in that they show a positive temperature dependence of the saturation, amounting between 1.39°K and 1.99°K to over 15%, and with those of Strauss (Long and Meyer 1953 a) in amount adsorbed. In view of these discrepancies between different measurements, even in the same laboratory, the disagreement of our results should be considered rather as in scale than in kind.

Agreement is even less satisfactory in the flow experiments on the unsaturated film. Using an arrangement similar to that introduced by Brown and Mendelsohn (1947) for separating the films from the gas phase, Long and Meyer (1952 a) find two entirely different sets of results when employing two different methods of measurement (called by them Method I and Method II). They bring reasons why Method II should be the one giving the correct answer. However, observations with a different arrangement (Bowers, Brewer, and Mendelsohn 1951), subsequently confirmed by Long and Meyer (1952 b), give results in agreement with Method I.

It thus appears that the attempt of Long and Meyer (1953 a, b) to express the condition of unsaturated helium below the lambda point by a simple physical adsorption becomes more difficult as further results come to hand. Indeed one feels that already any of the inconsistencies in their own results, if followed up, might have led these authors to conclusions similar to ours.

The confusion of results outlined above strongly indicates, in fact, that in helium II the surface phenomena, like those of heat conduction and viscosity in the bulk liquid, depend on the method of measurement. On the basis of our experiments, we have tentatively suggested a new effect—the formation of small clusters of bulk liquid below the saturation pressure—which owe their stability to zero point energy and, which will overlap and obscure the process of van der Waals adsorption. At this stage, when comparatively little about this new effect is known, it would be premature to try to determine the degree to which any of the existing adsorption measurements are falsified. Thus, in the present state of knowledge, detailed derivations based on adsorption theory, such as carried out by Long and Meyer (1953 a, b) may be somewhat too confident in the case of helium below the lambda point.

REFERENCES

BOWERS, R., BREWER, D. F., and MENDELSSOHN, K., 1951, *Phil. Mag.*, **42**, 1445.
 BREWER, D. F., and MENDELSSOHN, K., 1953 a, *Phil. Mag.*, **44**, 340 ; 1953 b, *Ibid.*, **44**, 559.
 BROWN, J. B., and MENDELSSOHN, K., 1947, *Nature, Lond.*, **160**, 670.
 FREDERIKSE, H. P. R., and GORTER, C. J., 1950, *Physica*, **16**, 402.
 LONG, E. A., and MEYER, L., 1949, *Phys. Rev.*, **76**, 440 ; 1950, *Ibid.*, **79**, 1031 ; 1952 a, *Ibid.*, **85**, 1030 ; 1952 b, *Ibid.*, **87**, 153 ; 1953 a, *Advances in Physics*, **2**, 1 ; 1953 b, *Phil. Mag.*, **44**, 788.
 STRAUSS, A. J., 1952, *Thesis*, Chicago.
 STRAUSS, A. J., MEYER, L., and LONG, E. A., 1951, *Proc. Int. Conf. on L. T. Physics* (Oxford), p. 91.

The Martensitic Transformation in Cobalt

By Z. S. BASINSKI and J. W. CHRISTIAN
 The Inorganic Chemistry Laboratory, Oxford

[Received May 18, 1953]

In a previous letter (Anantharaman and Christian 1952), it was stated that there is no simple mechanism for the production of the (macroscopically) homogeneous shear, required for the cobalt transformation. In fact, a suitable dislocation node may be constructed in a manner entirely analogous to that used by Thompson and Millard (1952). When a major dislocation emerges from a region of h.c.p. lattice, it may dissociate into allowed dislocations of the f.c.c. lattice in the following way :

$$c[00\cdot1] = \frac{a}{2}[110] + \frac{a}{2}[011] + \frac{a}{6}[1\bar{2}1].$$

The first two dislocations in the cubic lattice are perfect, and may lie in the $(1\bar{1}1)$ plane, in which they can both glide. The third dislocation, which moves in the (111) plane, is the transformation dislocation, and by rotating about the node, the h.c.p. region is extended by two atomic planes for each complete revolution. The node is anchored since the major dislocation cannot normally glide, and the mechanism is unaffected if the perfect cubic dislocations dissociate into $a/6<112>$ partial dislocations.

If we have two f.c.c. lattices in twinned orientation, an analogous node is formed from three dislocation lines, which are $a/2<110>$ type dislocations in the parent and twin lattices respectively, and the $a/6<112>$ twinning dislocation. Growth of the twin is then geometrically possible, but the node is not anchored since both the perfect dislocations may glide.

In an accompanying note Dr. B. A. Bilby points out that Millard and Thompson's method of constructing a node is perfectly general. In view

of Bilby's result, the example discussed here is trivial, but is published to correct the statement of the previous letter. The proposed mechanism does not explain how the h.c.p. nuclei originate.

We should like to thank Dr. Bilby for sending us details of his general theory, and Dr. W. Hume-Rothery, F.R.S. for his interest.

REFERENCES

ANANTHARAMAN, T. R., and CHRISTIAN, J. W., 1952, *Phil. Mag.*, **43**, 1338.
THOMPSON, N., and MILLARD, D. J., 1952, *Phil. Mag.*, **43**, 422.

LXXXIII. *Notices of New Books and Periodicals received*

Superconductivity. By D. SHOENBERG. [Pp. x+256.] (London : Cambridge University Press, 1952.) Price 30s. net.

THE first edition of Dr. Shoenberg's *Superconductivity* appeared in 1938 in the Cambridge Physical Tracts. The second edition now appears in the new series of Cambridge Monographs on Physics and presents a much enlarged and up to date account of the subject. The chapters on the magnetic and thermal properties of superconductors of macroscopic dimensions have been revised and somewhat extended to include thermal conductivity and thermoelectric effects. The major increase in the size of the book comes in the chapters on the intermediate state and the penetration of a magnetic field into superconductors. Here much recent work is discussed, particularly the remarkable experimental investigations of Meshkovsky and Shalnikov of the discontinuous structure of the intermediate state and the determination of the high-frequency resistance of superconductors. The last chapter gives an account of the phenomenological theory and of the more recent attempts at a fundamental theory of superconductivity.

The book is written from the point of view of the experimental physicist, the mathematics being restricted to that required for the concise statement of the argument. It may be warmly recommended as an excellent account of the present state of this rapidly growing subject. The inclusion of many graphs and tables of data make it a valuable reference manual for research workers.

L. C. J.

Stress Waves in Solids. By H. KOLSKY. [Pp. 211.] (Monographs on the Physics and Chemistry of Materials.) (Oxford : Clarendon Press, 1953.) Price 25s.

THE appearance of a book solely devoted to the propagation of stress waves in solids is very welcome. Hitherto, with the exception of Rayleigh's classic, the treatment accorded to the subject of wave propagation in perfectly elastic solids has not been ideal. In books on elasticity it has either been presented as a piece of dry mathematics or, where the engineering applications have been emphasized, it has been relegated to a very secondary position. Nowhere has wave propagation in imperfectly elastic solids been adequately discussed ; now that the necessary experimental techniques have been developed, applications to fundamental studies of internal friction in metals, and the behaviour of organic polymers at high rates of strain, are becoming increasingly numerous.

Dr. Kolsky has undertaken the writing of a concise account of this new field, essentially from a physical and experimental standpoint. It has, however, been necessary to draw fairly heavily on the classical theory of elastic waves and steady-state vibrations, and nearly one half of the book is concerned with this. Waves in visco-elastic and perfectly plastic solids are also treated, though relatively briefly and with rather few references to original papers. Elsewhere the bibliography is entirely adequate.

One minor stylistic blemish occurs repeatedly in the first half of the book, though curiously enough not in the second. This is the interpolation of a colon before an equation, irrespective of whether it is grammatically necessary. This is, unfortunately, a not uncommon fault in present-day scientific literature.

The task of selection from so many diverse sources must have been a difficult one. It is inevitable that different readers will find one or other section too short ; for the reviewer, it is the chapter on the role of stress waves in fracture. Individual tastes apart, Dr. Kolsky has succeeded admirably, and has written clearly and interestingly throughout.

R. H.

Vacuum Technique. By A. L. RIEMANN. (Chapman and Hall.) [Pp. 430.] Price 50s.

IT is apparent that the author has made a wide and careful study of the books and literature already in existence relating to the subject. This knowledge he has brought up to date with the latest information on pumps, materials and the various processes allied with vacuum work, and full references have been given should the reader desire to study the original papers. To this he has added a large amount of detailed practical information obviously derived from personal experimental experience and close contact with valve manufacture, so that the information contained in certain sections such as Metal Glass Sealing and, in particular, Copper Glass Seals, is unusually comprehensive. There are few points which have not been discussed, and occasionally where an aspect such as general glass working has not been fully dealt with, excellent references have been given. Where desirable, sufficient theory has also been incorporated, together with extensive statistical data. The presentation is in a form likely to be most useful to the student or technician actually engaged on vacuum work, who will find in this book adequate information on pumps, gauges, getters, materials and processes that are in current use. J. H. B.

Progress in Nuclear Physics. Vol. 2. Edited by O. R. FRISCH. (London : Pergamon Press Ltd.) [Pp. 295.] Price 63s.

Annual Review of Nuclear Science. Vol. 2. (Annual Reviews Inc., Stanford, California.) [Pp. 429.] Price \$6.00.

THE rapid development of nuclear physics, and the increasing segregation of the subject into separate branches, are made very clear by the contents of these two volumes. Each consists of a number of separate articles on different fields, in which the reader is taken to the limit of knowledge at the time of writing, and offered an extensive bibliography in case he wishes to study the field further or to obtain more detailed information.

The articles in Professor Frisch's book are, in his own words, intended to help both the nuclear physicist in finding information in fields adjacent to his own, and the other scientist in getting an introduction to some technique in nuclear physics which he may wish to use. Accordingly, they are clearly written, with adequate introduction, illustration and discussion, so that each might be considered as a miniature text-book. There are some cross-references between the articles, and the coordinating hand of the editor is occasionally evident. The individual articles deal with : magnetic β -ray spectrometers ; nuclear paramagnetic resonance ; luminescent materials for scintillation counters ; the neutron-proton interaction ; fission ; the low-lying excited states of light nuclei ; the nuclear shell model ; ionization by fast particles.

The American review contains articles of a slightly different character : they tend to be directed more towards the specialist, and to be presented in a way which makes them less easily understood by Professor Frisch's "other scientist". The articles are concerned with : the origin and abundance of the elements ; energy production in stars ; natural radiocarbon ; accelerators ; nuclear reactions induced by high-energy particles ; radiation effects in solids ; isotopes ; nuclear moments ; β -decay ; the origin and propagation of cosmic rays ; nucleon-nucleon scattering ; high energy fission.

The chief general criticism of these two volumes is that neither contains any attempt to review the state of nuclear physics as a whole ; in the reviewer's opinion, each title would be made more appropriate and each book improved by the addition of a chapter or introduction in which some such attempt was made, briefly and in general terms. Nevertheless, the usefulness of both books is beyond doubt.

W. M. G.

Micrometeorology. A Study of the Physical Processes in the Lowest Layers of the Earth's Atmosphere. By O. G. SUTTON. (McGraw-Hill, 1953.) Price 61s.

THIS book, by one of the best known authorities on atmospheric turbulence, fills an important gap in meteorological literature. For, apart from the short monograph on the subject written by Professor Sutton a few years ago, there has been no text-book in the English language entirely devoted to the theory of the physical processes in the surface layers of the atmosphere. The main title of the new book is too broad to indicate its scope and agricultural meteorologists might expect to find more about microclimatology included. Also it might be argued that some of the problems with which the book is concerned, e.g. the approach to the geostrophic wind and evaporation from the oceans, are outside the micrometeorological field. However, it would be difficult to draw a hard and fast line between micro- and macro-meteorological phenomena and in this book the sub-title and preface make the author's intentions quite clear.

Professor Sutton explains that the regions of the atmosphere with which he is concerned are those where life is most abundant. Perhaps we should refer to these lowest layers of the troposphere as the ' biosphere '. They are the layers that are close enough to the ground to be affected more by the surface itself, both directly and indirectly (for example, through its effect on air temperature and water content) than by the meteorological effects of the earth's rotation. The phenomena that the book deals with include laminar and turbulent motion in the lower atmosphere, the diffusion of momentum, matter and heat, the effect of radiation, and wind and temperature structure. Professor Sutton brings together the essentials of all the more important researches on these problems which have been published during the past 30 years or so, but he includes the basic physics necessary for a proper understanding of the subject and the book is, therefore, self-contained. Each chapter starts with a brief survey of the subject, which is perhaps a good alternative to concluding each chapter with a summary of the important points. Professor Sutton writes as a mathematical physicist and his treatment is essentially theoretical. He does not describe experimental techniques in any great detail but he does discuss the practical applications of the theoretical work.

One can hardly assess the full value of this book from a first reading ; a book of this type must be used as a working tool in a particular problem before its true worth can be properly appreciated. For anyone undertaking investigations involving atmospheric phenomena near the earth's surface Professor Sutton's book undoubtedly provides an excellent starting point ; it relieves the investigator of the trouble of searching through a wide field of published papers.

The book is well produced (but there are one or two omissions in the name index and a few misprints) and like most McGraw-Hill publications it is very strongly bound.

F. J. S.

NOTES FOR THE GUIDANCE OF AUTHORS SUBMITTING PAPERS FOR PUBLICATION.

Papers should be in typescript with double spacing. One side only of the paper should be used. MSS. should be as brief as is consistent with clarity. In particular the citation of elementary steps in a mathematical argument is to be avoided.

An abstract should always be provided. This should be as informative as possible, should be placed at the head of the paper and should not exceed 200 words in length.

USE OF CERTAIN MATHEMATICAL SYMBOLS.

In mathematical expressions appearing in the solid text the task of the compositor can be appreciably lightened in several ways, such as the use of the solidus and of the *exp* notation with a careful employment of the bracket. Thus, write $\sin(\theta/2)$ not $\sin\left(\frac{\theta}{2}\right)$; $(\sin\theta)/2$ not $\frac{\sin\theta}{2}$; $(a+b)/(c+d)$

not $\frac{a+b}{c+d}$; $a+b/c+d$ not $a+\frac{b}{c}+d$. Also $\sqrt{(a^2+b^2)}$ or $(a^2+b^2)^{\frac{1}{2}}$ not

$\sqrt{a^2+b^2}$; $n!$ not $|n|$; and $3kT/2$ not $3/2kT$ —a common error. An example of the *exp* notation is Andrade's equation: $\eta=A[\exp(B/T)]$.

In a mathematical argument, however, the formulæ or equations should be written out on separate lines (*displayed* is the technical term) and numbered. The necessity for the use of the solidus is not so pressing.

ILLUSTRATIONS.

Line drawings should be made in Indian ink on Bristol board or tracing-cloth. Foolscap size should be the maximum, and unless the work is executed by a professional draughtsman the diagrams should be lettered lightly in pencil. Photographs are usually reproduced as plates, and should not be used unless absolutely necessary.

REFERENCES.

References should be made in the solid text by giving the author's name and year of publication in brackets. Thus: 'It has been shown (Jones 1935) that' If Jones has published more than one paper in 1935, the papers should be referred to in order of date as (Jones 1935 a, b, . . .). References should be collected in alphabetical order of authors and placed at the end of the paper under the heading 'References'. Each reference should be of the form: Author's name; year of publication; abbreviated title of journal; series (if any) in square brackets; volume in Clarendon arabic type; page number. Thus: Jones, A. B., 1935, *Phil. Mag.* [7], **19**, 742.

ABBREVIATIONS.

The Royal Society recommends the following notation for multiples and submultiples of any unit:

10^6	10^3	1	10^{-1}	10^{-2}	10^{-3}	10^{-6}	10^{-9}	10^{-12}
M	k		d	c	m	μ	μm	$\mu\mu$

Further information may be obtained from:

- (1) The Royal Society's pamphlet 'Notices on the preparation of papers communicated to the Royal Society';
- (2) Report by the Symbols Committee of the Royal Society, 1951. Price 9d. per copy. The Royal Society, Burlington House, London, W.1.